STORAGE-ITEM
MAIN

LP9-L24G
U.B.C. LIBRARY

THE LIBRARY

THE UNIVERSITY OF
BRITISH COLUMBIA

THE INTERNATIONAL UNIVERSITY SERIES IN PSYCHOLOGY

Edited by

CARL MURCHISON, Ph.D.

Professor of Psychology and Director of the
Psychological Laboratories in Clark University

THE INTERNATIONAL UNIVERSITY SERIES IN PSYCHOLOGY

PSYCHOLOGIES OF 1925

By Madison Bentley, Knight Dunlap, Walter S. Hunter, Kurt Koffka, Wolfgang Köhler, William McDougall, Morton Prince, John B. Watson, and Robert S. Woodworth. *Edited by* Carl Murchison.

CRIMINAL INTELLIGENCE

By Carl Murchison, Ph.D., *Professor of Psychology and Director of the Psychological Laboratories in Clark University.*

THE CASE FOR AND AGAINST PSYCHICAL BELIEF

By Sir Oliver Lodge, Sir Arthur Conan Doyle, Frederick Bligh Bond, L. R. G. Crandon, Mary Austin, Margaret Deland, William McDougall, Hans Driesch, Walter Franklin Prince, F. C. S. Schiller, John E. Coover, Gardner Murphy, Joseph Jastrow, and Harry Houdini. *Edited by* Carl Murchison.

FEELINGS AND EMOTIONS: THE WITTENBERG SYMPOSIUM

By A. Adler, F. Aveling, V. M. Bekhterev, M. Bentley, G. S. Brett, K. Bühler, W. B. Cannon, H. A. Carr, Ed. Claparède, K. Dunlap, R. H. Gault, D. W. Gruehn, L. B. Hoisington, D. T. Howard, E. Jaensch, P. Janet, J. Jastrow, C. Jörgensen, D. Katz, F. Kiesow, F. Krueger, H. S. Langfeld, W. McDougall, H. Piéron, W. B. Pillsbury, M. Prince, C. E. Seashore, C. E. Spearman, W. Stern, G. M. Stratton, J. S. Terry, M. F. Washburn, A. P. Weiss, and R. S. Woodworth. *Edited by* Martin L. Reymert.

SOCIAL PSYCHOLOGY: THE PSYCHOLOGY OF POLITICAL DOMINATION

By Carl Murchison, Ph.D., *Professor of Psychology and Director of the Psychological Laboratories in Clark University.*

THE COMMON SENSE OF DREAMS

By Henry J. Watt, D.Phil., *Late Lecturer in Psychology in the University of Glasgow, and Late Consulting Psychologist to the Glasgow Royal Asylum. Author of "The Psychology of Sound."*

THE FOUNDATIONS OF EXPERIMENTAL PSYCHOLOGY

By H. Banister, Philip Bard, W. B. Cannon, W. J. Crozier, Alexander Forbes, Shepherd Ivory Franz, Frank N. Freeman, Arnold Gesell, H. Hartridge, Selig Hecht, James Quinter Holsopple, Walter S. Hunter, Truman L. Kelley, Carney Landis, K. S. Lashley, Mark A. May, T. H. Morgan, John Paul Nafe, George H. Parker, Rudolf Pintner, Eugene Shen, L. T. Troland, and Clark Wissler. *Edited by* Carl Murchison.

THE PSYCHOLOGICAL REGISTER

Edited by Carl Murchison, *Clark University,* in cooperation with F. C. Bartlett, *University of Cambridge,* Stefan Blachowski, *University of Poznan,* Karl Bühler, *University of Vienna,* Sante De Sanctis, *University of Rome,* Thorleif G. Hegge, *University of Oslo,* Matataro Matsumoto, *Tokyo Imperial University,* Henri Piéron, *University of Paris,* and A. L. Schniermann, *Bekhterev Reflexological Institute.*

THE FOUNDATIONS OF EXPERIMENTAL PSYCHOLOGY

THE INTERNATIONAL UNIVERSITY SERIES IN PSYCHOLOGY

THE FOUNDATIONS OF EXPERIMENTAL PSYCHOLOGY

by

H. Banister
Philip Bard
W. B. Cannon
W. J. Crozier
Alexander Forbes
Shepherd Ivory Franz
Frank N. Freeman
Arnold Gesell
H. Hartridge
Selig Hecht
James Quinter Holsopple
Walter S. Hunter
Truman L. Kelley
Carney Landis
K. S. Lashley
Mark A. May
T. H. Morgan
John Paul Nafe
George H. Parker
Rudolf Pintner
Eugene Shen
L. T. Troland
Clark Wissler

Edited by
Carl Murchison

WORCESTER, MASSACHUSETTS
CLARK UNIVERSITY PRESS
LONDON: HUMPHREY MILFORD: OXFORD UNIVERSITY PRESS
1929

Copyright, 1929, by
Clark University
all rights reserved

Second Printing, November, 1930

BF
191
M8

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

It has not been easy to determine the chapters that should appear in this book. The list as it appears has been decided upon after consultation with many who are interested in this type of approach to the problems of psychology. It is not likely that any fruitful problem has been omitted. Those problems that have been omitted are thereby classified as promising little return from experimental effort. I know no better way of clearing the débris from the intellectual decks of psychology. It has too long been the custom to discuss over and over every problem that had ever been discussed before, as if some value had accrued from mere discussion. The only criterion of inclusion that has been considered during the preparation of this volume is whether a given problem promises a reward to experimental investigation.

It is our intention that this book shall appear in new editions frequently enough to be always abreast of progress. That will mean a revision almost every year. It is unthinkable that an important exposition of the problems of experimental psychology ever be in any other condition.

As time passes and emphasis in psychology changes, it may be expected that there will be revision of the list of problems given exposition in succeeding editions of this book.

I wish to express once more my grateful thanks to Dr. Walter B. Cannon for his generosity and personal sacrifice in preparing the chapter on hunger and thirst in a few days' time because of the illness of the one whose authorship of this chapter had been announced. I wish also to express my warmest gratitude to Dr. A. Navez, who did much detailed work for Dr. W. J. Crozier during the latter's illness.

It has been immensely interesting to have been associated with the making of this book. I have not yet found anything more intriguing or promising more. We hope that friends of psychology will not hesitate to tell us everything we ought to know in our further cooperative efforts to make a standard, comprehensive, and progressing guide to the fruitful problems of psychology.

CARL MURCHISON

CLARK UNIVERSITY
WORCESTER, MASSACHUSETTS
July 23, 1929

TABLE OF CONTENTS

Preface vii

1. The Mechanism and Laws of Heredity 1
T. H. Morgan, California Institute of Technology

2. The Study of Living Organisms 45
W. J. Crozier, Harvard University

3. The Mechanism of Reaction 128
Alexander Forbes, Harvard Medical School

4. Vision: I. Visual Phenomena and Their Stimulus Correlations 169
L. T. Troland, Harvard University

5. Vision: II. The Nature of the Photoreceptor Process . . 216
Selig Hecht, Columbia University

6. Hearing I 273
H. Banister, Cambridge University

7. Hearing II 313
H. Hartridge, St. Bartholomew's Medical School
H. Banister, Cambridge University

8. The Chemical Senses 350
George H. Parker, Harvard University
W. J. Crozier, Harvard University

9. The Sense of Feeling 392
John Paul Nafe, Clark University

10. Space and the Non-auditory Labyrinth 414
James Quinter Holsopple, formerly of Yale University and of Western Reserve University

11. Hunger and Thirst 434
W. B. Cannon, Harvard Medical School

12. Emotion: I. The Neuro-humoral Basis of Emotional Reactions 449
Philip Bard, Princeton University

13. Emotion: II. The Expressions of Emotion 488
Carney Landis, Wesleyan University

14. Learning: I. Nervous Mechanisms in Learning . . . 524
K. S. Lashley, Behavior Research Fund, Chicago

15. Learning: II. Experimental Studies of Learning . . . 564
Walter S. Hunter, Clark University

16. The Individual in Infancy 628
Arnold Gesell, Yale University

17. The Individual in School: I. General Ability . . . 661
Rudolf Pintner, Teachers College, Columbia University

18. The Individual in School: II. Special Abilities and Their Measurement 705
Frank N. Freeman, University of Chicago

19. The Adult in the Community 738
Mark A. May, Yale University

20. The Conflict and Survival of Cultures 786
Clark Wissler, Yale University

21. The Abnormal Individual 809
Shepherd Ivory Franz, University of California at Los Angeles

22. General Statistical Principles 832
Truman L. Kelley, Stanford University
Eugene Shen, China Institute in America

23. The Statistical Treatment of Certain Typical Problems . 855
Truman L. Kelley, Stanford University
Eugene Shen, China Institute in America

Name Index 884

Subject Index 893

CHAPTER 1

THE MECHANISM AND LAWS OF HEREDITY

T. H. MORGAN
California Institute of Technology

If newly acquired habits of an individual should cause corresponding structural and psychic changes in its offspring, heredity would become a branch of animal psychology. The slogan that instincts are inherited habits has, it is true, at times held the attention of philosophers, and even scientists of the old school. Lamarck, the acknowledged founder of the doctrine of the inheritance of acquired characters, brought forward no sufficient evidence in support of his views, since the cases he cites by way of illustration—the giraffe's long neck as a result of stretching, the toe membranes of wading birds from spreading the toes in mud, etc.—are largely figments of the imagination. In fact, the idea itself did not originate with Lamarck, but is a very ancient one. A belief in such transmission is found in the folklore of many peoples, and Lamarck's "evidence" is indeed as naïve as that of these stories and myths. It is, I think, largely because Lamarck relied on argument rather than objective proof that his theory failed to be widely accepted by physiologists trained in the more exact methods of science, and found its adherents largely amongst artists, amateurs, paleontologists, psychologists, and zoölogists not addicted to the rigors of experimental science.

Herbert Spencer (186-191), at the end of the last century, carried the message to a large audience, ready to believe but not over-informed as to the difficulties inherent in the problem. Spencer brought forward no new evidence but exploited to full advantage certain theoretical situations. Perhaps his most characteristic argument relates to the discrimination of the sense of touch in different parts of the human body. A study of the topography of this function shows, he claimed, that this sense is most refined wherever the skin is most exposed to contact with the outside world —on the lips, fingertips, etc.—and decreases in proportion to the distance from the extremities in question. The discrimination rests on the distribution of touch-buds, but Spencer failed to show that these would develop more abundantly in an individual where contact is most frequent, and more abundantly in the offspring of the individual in question. Here, as in all such cases, the real point at issue is assumed on the basis of circumstantial evidence and the biological difficulties ignored. Moreover, when the theory does not strictly apply, imaginary possibilities are introduced; and where it is the simplest solution, no actual evidence is supplied to establish it, but its plausibility is urged as sufficient argument for its validity.

Hering (66) in 1870 pointed out similarities between memory and heredity and because of a striking but superficial resemblance attempted to identify one with the other. The argument like others of its kind ignores the actualities in the two situations. These differences have become greater as our information in the two fields has become more specific. Samuel Butler (17, 18), better known as a romantic novelist and writer than as a scientist, wrote vehemently in behalf of the identification of memory with heredity, undaunted by the difficulty of attempting to explain the better-known phenomena of heredity by an appeal to the less well-understood processes of memory. Somewhat similar attempts in later years have been made by Ward (198*a*), Semon (172-176), and Eldridge (39), not to mention less ambitious attempts by others.

More serious consideration must be given to the experimental evidence recently brought forward by Griffith (59) and later reviewed by Detlefsen (33, 34), also to the careful experiments of McDougall (95) on the inherited effect of training of white rats.

Griffith's experiments (59) with continuous rotation appear to furnish a case of inheritance of specific disequilibration. Adult rats were rotated in pens for long periods, some in a clockwise and others in a counterclockwise direction. Several weeks after removal some of the rats developed a peculiar twist of the head and a permanent nystagmus in direct relation to the direction of rotation, etc. Griffith also reported the appearance of disequilibration in the descendants of these rats, and, what is more important, in the earlier oral reports by Detlefsen and in Griffith's report, there was stated to be a high correlation between the right or left reaction of the treated parents and that of their offspring. Later observations by Cassamajor, Detlefsen (33), McCordock and Congdon (94), have shown that disequilibrated rats occasionally appear without previous rotation. These rats give the same physiological reaction that the rotated rats exhibit. There is more than a suspicion that the reaction is due to defects or disease in one of the ears. Detlefsen has also called forth the same reaction by injuring one labyrinth. It seems, therefore, not improbable that the rotation caused similar injuries, but this still will not explain the reported correlation between parent and offspring in Griffith's experiment. The not infrequent occurrence, however, of disequilibrated rats that have not been rotated cannot but jeopardize the results of Griffith's experiments, even although he found no such disequilibration in the brothers and sisters of his treated stock or in their descendants. The experiment calls for careful repetition in the light of this evidence on rats that have been made as nearly homozygous as possible beforehand, and also for careful scrutiny as to possible infection or indirect transmission of a diseased condition of the ears.

McDougall's (95) elaborately planned experiments with white rats dropped into water and given a choice of routes for escape with or without punishment seem to him to prove that the results of training were in some degree inherited. Until the experiments are completed, it is hazardous to criticize them on the basis of the published data, but in so far as reported the results suggest the following questions. The apparent implication in the statement that, since the rats had long been inbred, the stock was thoroughly homogeneous, is offset by another statement that there were "very large individual differences between the animals in respect of the rate at which they master their tasks." In another connection McDougall points out that some of the rats and their descendants showed a tendency to go to the right (or left). This would lead a geneticist to question the suitability of this or closely related stock for this particular purpose. It is disturbing, to say the least, to meet with the statement, "there is some indication that the transmission of effects of training follows a law of segregation reminiscent of Mendel's law, the effects being transmitted to some individuals and not at all to others." The evidence for these statements is too vague to mean much, if anything. The probable errors, when it is possible to calculate them from the data, leave the reader in doubt as to the value of the evidence in several cases. The participation of several experimenters (McDougall and students) in some of the experiments introduces another unknown quantity. In the absence of published data to indicate how long after training the young of the next litter were born, a serious question arises as to whether the effects of the training may or may not be lost. Is habit formation an irreversible process? If a new habit can be "formed" in 15 days, how much, one asks, may be lost after several months? Does the "loss" or "disuse" occur also in the germ-cells of the individual, or does the registration once made there last forever? It is surprising that these considerations should be omitted from a program as elaborate as this one of McDougall's. He avoids them perhaps by saying that his first endeavor was to determine the facts of inheritance; but, granting this, why should a psychologist not consider the possible effect of the waning of a habit, and its influence on the germ-cells? To be sure this complicates the problem, but I think, if such acquirements as these are really inherited, we must be prepared either to face complications or declare a miracle. However, in the absence of any knowledge as to how much training could possibly affect the germ-cells (Cimmerian darkness as black as McDougall's phrase for Darwin's theory of selection), this point need not be further discussed. McDougall himself raises the question as to what was inherited, whether an increase of general intelligence, or strength of the fear instinct, or a tendency to avoid the brightly illuminated route. He admits that the data do not throw any light on this question.

Physiologists can show a small but no less distinguished list of advocates of the doctrine. The experiments of Brown-Séquard on guinea pigs have been for a long time in the literature, but owing to the conflicting and largely negative evidence that has resulted from attempts to repeat his experiment his case has slowly drifted into the background. Romanes (159) has discussed this evidence at length, and more recently it has been considered by several other observers.

Brown-Séquard (9-16) reported the appearance of epilepsy in the offspring of parents that had been made epileptic by an injury to the spinal cord or to the sciatic nerve. He also reported partial closure of the eyes in animals born of parents showing a similar state of the eyelids after sectioning of the superior cervical ganglion. Exophthalmia appeared in the offspring of parents in which injury to the restiform body had produced a protrusion of the eyeball. Absence of some of the toes occurred in the offspring of animals whose parents had eaten off some of their toes, or in which the toes had been affected after cutting the sciatic nerve. These and a few other observations seem to relate to cases of specific inheritance. Brown-Séquard states that he had never observed similar defects in guinea pigs not operated upon. Many attempts to confirm these results have been made. A few cases of partial confirmation have been reported, as well as failures. Weismann's attempt to explain away the results as due to direct infection of the offspring by the parent will not account for the reported specific nature of the transmission. The most serious criticism of Brown-Séquard's results is the sporadic occurrence of the reported transmission, the absence of data to show the frequency of failure to transmit, and the failure to use pedigree material known to be free from inherited defects of the kind reported. In the case of epilepsy the nature of the disease (?) is problematical. The suggestion that it may appear in animals weakened by treatment or otherwise has been regarded as plausible. Whether the offspring of animals which have been made epileptic produce weaker offspring is not known. The absence of information on these and other conditions is a serious obstacle to the acceptance of Brown-Séquard's conclusions. Other experimenters have in recent years reported the appearance of defects in the offspring of guinea pigs whose parents have been injured, but it appears that these facts may be more simply explained as due to direct injuries to the germ-cells. Until such an explanation of Brown-Séquard's experiments can be shown not to hold, his results cannot be regarded as establishing his claims of somatic induction of specific defects.

A few years ago Pavlov startled the scientific world by a report of experiments made with mice. He stated that conditioned reflexes, established by the sound of a bell as a call for food, were transmitted to the offspring. Expectation hung for several years in the balance, for the great reputation of Pavlov called for serious consideration of his very positive statements based supposedly on exact experiments. Others who had already worked in the same field [MacDowell (90-92), Vicari (198), and Tolman] had seen no such effects where they should have been even more in evidence did they occur, since these workers dealt with the histories of individual cases and in one instance with homogeneous material. It now appears that some error had crept into Pavlov's experiments, since he has sponsored their withdrawal.

Zoölogists, too, can supply a long list of devoted adherents to the theory of the inheritance of acquired characters, and while Weismann's (200) attack in 1886 convinced almost everyone that more care was required in weighing the reported evidence, nevertheless there have not been lacking serious attempts to rehabilitate the theory on the grounds of new evidence.

We may pass over the views of the paleontologists, who could not possibly from the nature of their material supply the necessary critical evidence, for, however convinced they are that the course of evolution has been along lines of adaptation to environmental needs in a mechanical sense, they cannot be expected to disprove the possibility that such adaptations might have come about in other ways.

Zoölogists who have argued the question on general grounds, as did Darwin, and later Delage, Pauly (134), Semon (172-176), Cunningham (23-26), and others, have supplied no really critical evidence. On the other hand, the experimental work of several other zoölogists has kept the discussion alive. Kammerer's (76, 77) numerous reports of extensive series of experiments, claiming to establish the transmission of artificially induced structural changes and physiological activities, have been under fire from the beginning. The experiments were ingeniously planned, and taken at their face value appeared over-abundantly to establish his contention. To demonstrate his claims careful and unselected data are called for that have quantitative measurements to support them. In the absence of such data the expressed conviction of the author that such effects do occur introduces a personal equation that may well color the facts as reported. When his results have been repeated by others, with more attention to detail, they have failed to confirm his claims, and have made it clear that in his haste to establish his principles he has often neglected facts that vitiate the evidence. His final and most heralded result on the thumb pad of the midwife toad has been shown, in one crucial case at least, to have been fraudulent, the responsibility for which Kammerer repudiated before his tragic death.

Guyer and Smith (64, 65) reported a few cases in which eye defects in rabbits, induced *in utero,* reappeared in a few of the descendants. Later a more specific result was obtained by Guyer (62, 63) by injuring the lens with a needle. Again, eye defects were found in a few of the offspring. In criticism of this evidence it should be pointed out that the effects were at best sporadic and not beyond the possibility that they were due to defective germ-plasm in the stock used—the controls not being sufficient protection against this possibility. Guyer has been careful to guard against this contingency and has never made extravagant claims as to the certainty of the results being due to transmission. In more recent years several repetitions of the experiments [Huxley and Carr-Saunders (69), Findlay (45*a*), and others] have failed to confirm Guyer's results, and others [von Hippel, von Szily (196), Hochstetter, and Koyanagi] have reported eye defects in rabbits that occasionally crop up often enough to make the material suspect.

Certain moths have been reported to develop a black stain in certain factory districts in England. Harrison has made some experiments with these moths by putting into the food of the caterpillars some of the ingredients that might be supposed to have fallen on the larvae from factory smoke and to have caused the observed changes. The caterpillars so treated produced moths. A very few of the eggs of these moths gave melanic individuals and this character was transmitted to their offspring as a Mendelian character. In his earlier reports Harrison explained this as due to the inheritance of an acquired character in the conventional sense. A closer scrutiny of the evidence, however, does not support this conclusion, since it becomes evident that the result can be explained, and better explained, I think, as the result of a direct action on the germ-cells. Interpreted in this way, the body cells have not changed the germ-cells, at first indifferent, into their own likeness, so to speak, but a change took place in the preceding generation in some of the germ-cells that was transmitted both to the body cells and to the germ-cells of the black individuals. In other words, there was a direct action of the environment producing a mutation in one or more germ-cells and not "somatic induction." Dürkin's experiment purporting to show the inherited influence of light on pupae of a butterfly can be accounted for by the selection of individuals more or less sensitive to light from a mixed population.

That germ-cells may be affected by the environment has been shown by recent work on the effects of radium and X-rays [Little and Bagg (87), Muller]. The action of radium in producing irregularities in the distribution of chromosomes and parts of chromosomes had already been demonstrated by the earlier work of O. Hertwig (67, 68), Packard (131, 132), Mohr (99), and several others. Apparently in line with this evidence is that obtained by Stockard (193-195) on guinea pigs. The results were at first obscured by the fact that the lenses of the eyes of the adult guinea pig that had received alcohol treatment were made opaque. The offspring occasionally showed defects, and amongst these were eye defects of various kinds. Stockard is inclined to interpret the results as due to the direct effects of the alcohol on the germ-cells of the treated individuals.

An impartial survey of all this evidence cannot fail, I think, to give the impression that the many attempts to obtain crucial evidence that somatically acquired characters produce specific changes of the same kind in the germ-cells have failed. If such effects are transmitted, and especially if these are pre-eminently adaptive ones, there is every reason to suppose that evidence of such transmission should be obtainable—in fact should be abundantly apparent to anyone making careful observations. There is lack of care in experimentation and of acumen in reasoning from the evidence manifest in most of the reported results. To suggest that the influence may not be felt unless extended over many, or an indefinite number of generations, removes the problem from the field of experimental science. The situation is even worse when behavior patterns are considered, for all such adaptive reactions are highly complex, involving intricate nerve and muscle adjustments. It is, I venture to think, beyond the imaginative capacity of biologists, familiar with what is best known today in embryonic development and in genetics, to suggest any rational process by which such complexities could be carried over *as a whole* from the body cells to the germ-cells. If, then, the alliance between psychology and genetics is not likely to be established along this line, it remains to discover whether the present status of heredity points to other possible relations that may exist between these two branches of science.

Mendelian Heredity

In the meantime, while argument and counter-argument have continued without apparently bringing any relief or real support to the Larmarckians, there has been remarkable progress in several other fields that must have an important influence on all questions concerning heredity. I refer to the study of the egg, its fertilization and development, to the results of cytological work bearing on the nature of the chromosomes, and to the extraordinary series of events that occur at the time of the maturation of the germ-cells. These discoveries have become such an integral part of the modern theory of heredity that to ignore them through prejudice or ignorance would be tantamount to a disregard of the most accurate information that is known relating to the mechanism of heredity. The details of this cytological work will be found in numerous texts now available. Only the most general conclusions need be stated here.

The evidence from genetics has shown that the chromosomes are the

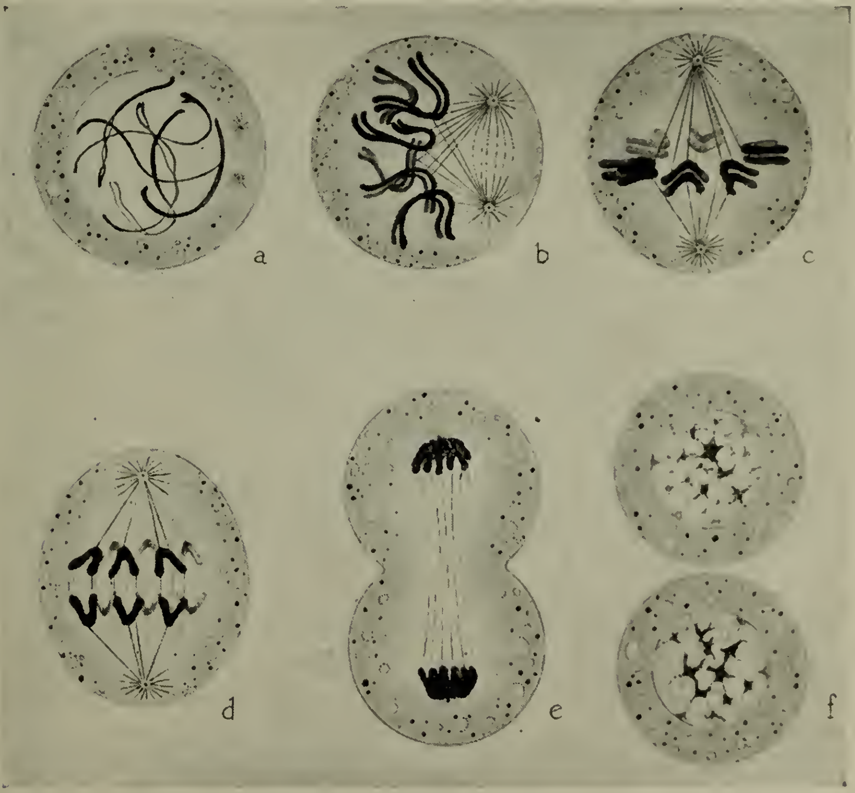


FIGURE 1

Diagram Illustrating Cell Division

The chromosomes are represented as black threads (*a*), or bent rods which become connected with protoplasmic threads (*b*). Each chromosome splits lengthwise (*c*), and the halves or daughter chromosomes pass to opposite poles (*d, e*), where they subsequently become vacuolated to form the resting nuclei of the two daughter cells (*f*).

bearers of the hereditary characters. Each species of animal and plant has a definite number of chromosomes of characteristic size. There are two chromosomes of each kind in all the body cells as well as in the young germ-cells. The members of each pair are, in homogeneous races, identical in their hereditary elements, but in the hybrid, or heterozygous individual, one member of a pair may differ from the other in one or more elements or genes. The cells are double-barreled or duplex with respect to their hereditary elements. The different pairs of chromosomes carry different sets of elements.

The genetic evidence further indicates that the hereditary elements, the genes, lie in a linear order in the chromosomes—like beads on a string—or at least the facts can be explained on this assumption, the genes themselves being too small to be visible. The serial order of the genes is maintained throughout all successive divisions of the cell and from one generation to the next. This is made possible by the way the chromosomes divide at each cell division. Each chromosome splits lengthwise into daughter threads, exactly equivalent (Figure 1). Since all the cells of the body contain the duplex set of genes, we must regard every cell, no matter how it is differentiated, as carrying the entire equipment of hereditary elements. In other words, the different cells of the body, no matter what their function, are looked upon as containing all the elements. It is difficult to furnish complete proof of the last statement, but since there is a great deal of direct evidence supporting it, and none opposed, it seems not unreasonable to make this somewhat broad generalization.

The implications in this statement are significant, for they mean that the embryonic development—the extensive processes of specialization or differentiation in the cells derived from the fertilized egg—cannot be explained, as Roux and Weismann once supposed, by the sorting out of the hereditary elements of the chromosomes, but must be explained on either of two hypotheses: first, that different genes begin to function at different stages of development; or, second, that the initial differences lie in the cytoplasm. In fact, the evidence at hand indicates that the initiation of differentiation of some of the important characters begins in the egg before the chromosomes are separated.

If differentiation of the parts of the individual is a property of the protoplasm of the egg and embryo, why, it may be asked, is not the protoplasm the real bearer of heredity? The answer is that there is today explicit evidence showing that the behavior of the cytoplasm is determined in the last analysis by the composition of the chromosomes. The determination may take place before development begins or at later stages. The evidence supporting this assumption will be given later, after the facts of Mendelian inheritance have been reviewed.

The Chromosome Mechanism of Heredity

Mendel's (96) two laws of heredity were deduced from numerical data derived from crosses of two individuals that differed in one or more characters. Since the types selected for crossing were sharply separable for one or more striking characters, these characters came later to be spoken of as unit characters.

Mendel assumed that each one of the contrasted characters is due to an element or unit in the germ-plasm responsible for that character, and that these units separate from each other in the eggs and pollen of the hybrid.

In the early days of Mendelism—at the beginning of this century—the idea prevailed in certain quarters that each element in the germ-cells produces in the individual one particular character. Needless confusion soon arose from the identification of the germinal unit with the unit character. The identification was misleading when it was attempted to show that the fluctuations in the character arose from similar fluctuations in the germinal unit. If this erroneous idea had gained headway it would have gone far to undo the all-important clue that Mendel's discovery had furnished as an explanation of heredity.

There is abundant evidence at the present time to prove that the germinal units or genes may and generally do affect more than one of the characters of the individual, and that these effects are present whenever a specific gene is present. As a rule the *most conspicuous* amongst these effects is designated as the character whose heredity we study. It would be equally possible in most cases to follow one of the minor characters. The outcome would be the same.

The idea that each character is the product of a single gene is in a sense correct, but only in so far as the presence of a particular gene is determinative for a given character. The genes are only differentials. But the same gene may affect many other characters; in fact, it might be said that a change in a pair of genes (or in one only if it is dominant) might affect every single character of the individual, but, since we cannot identify all of its effects, there is little to be gained in pressing this point. In fact, there are good and sufficient reasons to infer that some characters are not affected at all by certain specific genes, since "the principle of all or none" seems to be a rather general property of differentiation.

Conversely, if the genes frequently have such far-reaching influences as just suggested, it would seem to follow that a great many and perhaps all the genes are contributory agents to every character of the individual. The individual would be the effect of all of its genes in every part. It is not necessary at present to carry this statement too far, but it agrees with what embryonic development teaches with clarity, namely, that each organ passes through a series of changes before reaching its goal, and most organs are also composites of several or many kinds of tissues whose differentiation has taken place along different lines. The human eye, for example, comes in part from the forebrain, in part from the ectoderm of the side of the head. Mesenchyme and mesoderm contribute to it, and blood-vessels are also pushed into the interior of the mass. Therefore, when we contrast in heredity blue and brown eyes, we are describing a single differ-

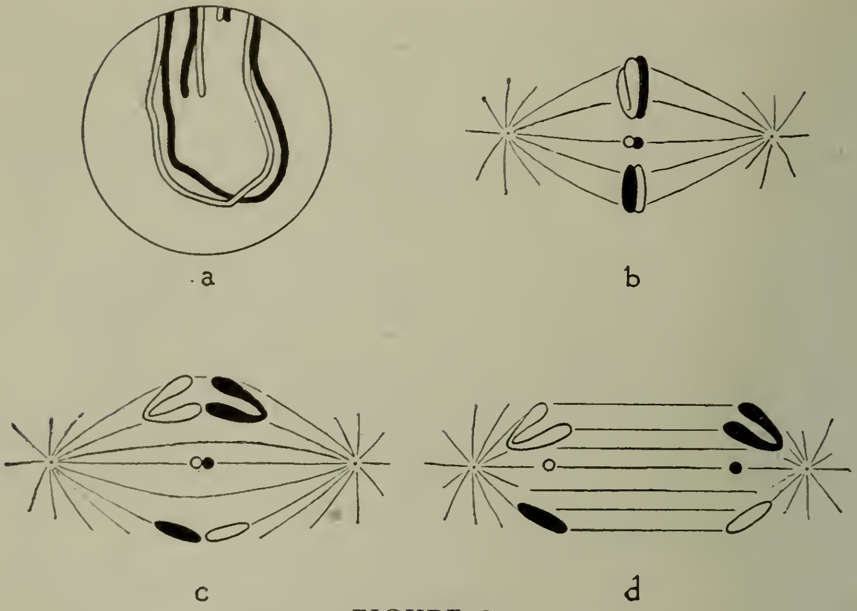


FIGURE 2

Diagram to Illustrate Conjugation and Reduction of the Chromosomes

In (*a*) pairs of chromosomes in the form of thin threads are coming together (conjugation). In (*b*) they have shortened and occupy the equator of the "spindle." In (*c*) one member of each pair is passing to one pole and its mate to the opposite pole of the spindle. In (*d*) they have reached the pole. Cell division then taking place, two cells result each with the half number of chromosomes. The second maturation division is not shown here.

ence dependent on the presence or absence of brown pigment granules in the iris (in which the two kinds of eyes differ), yet there remain numbers of characters that the eyes have in common. Here we pick out only a pair of contrasted conditions, but it may be pointed out that the same pair of genes that leads to a failure in the development of brown pigment may have other effects in other parts of the body—in the nervous system or in the skin, for instance—that we have not yet identified. In any case the eye is found to be a complex of tissues to whose development a host of genes may be contributory.

For present purposes, however, the important fact is that we can ascribe this effect on the eye color to a particular chromosome and in similar cases locate the gene in a limited region of a chromosome. In order to show in general how this has been possible, it will be necessary to describe how the mechanism that is implied but not expressed in Mendel's laws is due to the behavior of chromosomes as wholes, and how later work has enabled us to locate, within the chromosome chain of genes, the particular gene that produces certain effects.

Just before the final stages in the ripening of the germ-cells (egg and sperm) an extraordinary event takes place. In each cell the chromosomes come together in pairs. Each pair consists of one maternally derived and one paternally derived chromosome. This is illustrated in Figure 2*a*, where the maternal chromosomes are black and the paternal white.

At the time of conjugation, the chromosomes stretch out as thin threads —possibly a line of genes. They come together in pairs throughout their length and then shorten to form rods which are often bent.

The nuclear wall next disappears. The condensed chromosomes then pass onto a spindle that has meanwhile developed in the protoplasm (Figure 2*b*). Here they come to lie in a flat plate at the equator of the spindle. The members of each pair then pass to opposite poles (Figure 2 *c, d*). Around each set a new membrane develops.

Each daughter cell now contains one of each kind of chromosome. The original number is reduced to half in each daughter cell, and, since at the

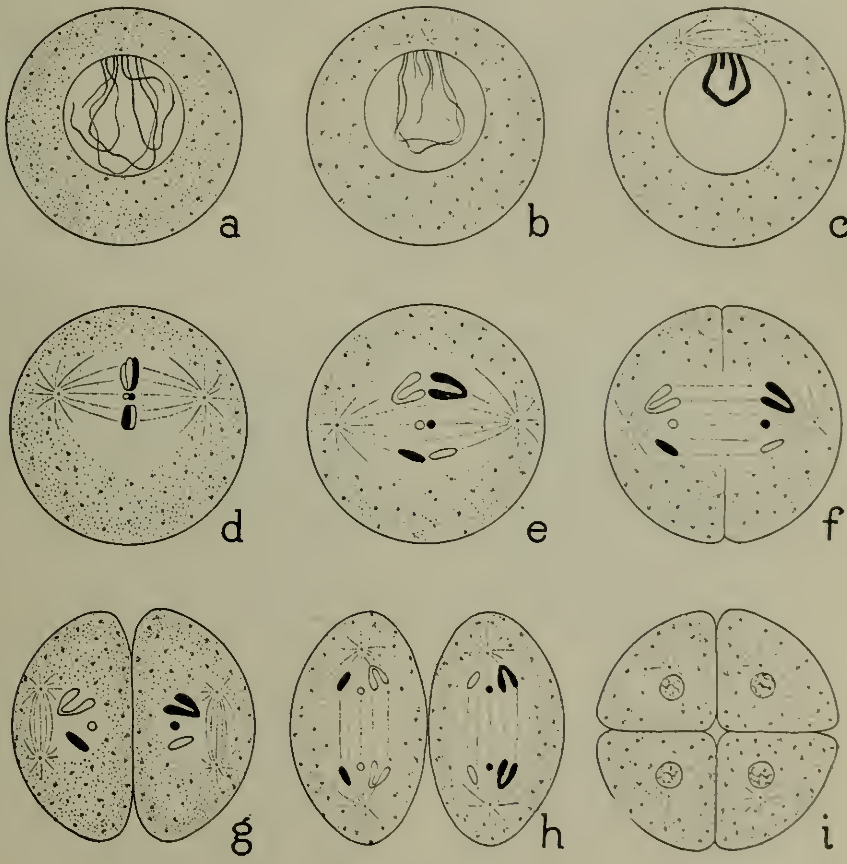


FIGURE 3

Diagram to Illustrate the Two Maturation Divisions of a Sperm Mother Cell

In the upper line (*a, b, c*) conjugation is shown. In the second line (*d, e, f*) the reduction at the first division is taking place. In the third line (*g, h, i*) the second or equatorial division is taking place in each of the two cells. In the latter division each chromosome splits lengthwise as in ordinary cell division.

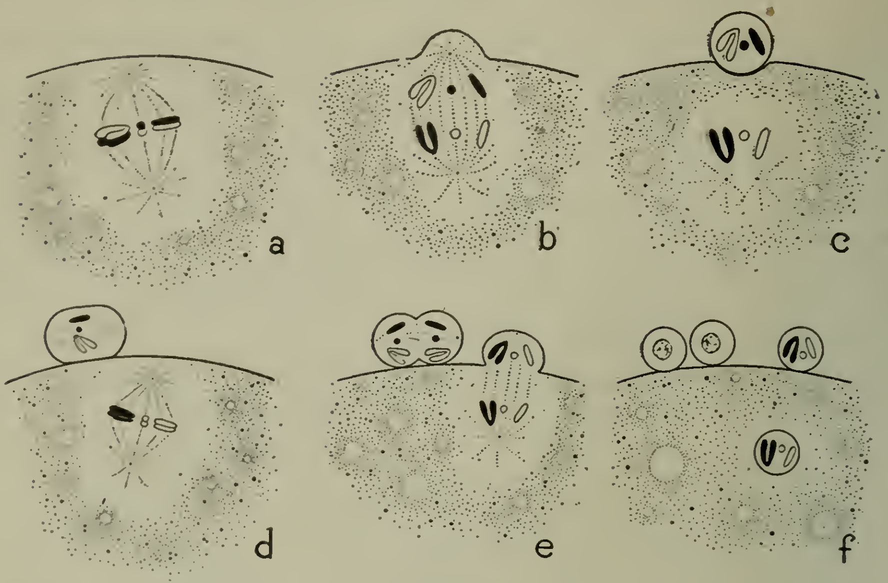


FIGURE 4

DIAGRAM TO ILLUSTRATE THE TWO MATURATION DIVISIONS OF THE EGG

In the upper line (*a, b, c*) the reduction division is taking place. The first polar body is given off (*c*). In the second line (*d, e, f*) the second or equatorial division is taking place, each chromosome splitting lengthwise. A second polar body is given off (*e, f*), and at the same time the first polar body divides.

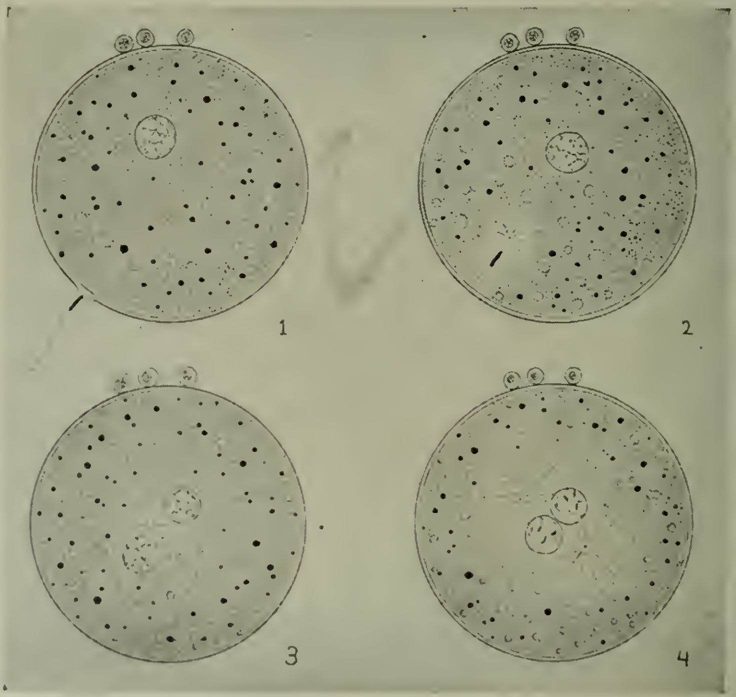


FIGURE 5

DIAGRAM TO ILLUSTRATE THE FERTILIZATION OF THE EGG BY A SINGLE SPERM WHICH BRINGS IN THE SPERM NUCLEUS

The two nuclei fuse to form the segmentation nucleus which will have the total number of chromosomes.

time of separation the maternal and paternal members of the different pairs are distributed at random, the half number may be made up both of maternal and paternal chromosomes, but each cell always has one member of each pair.

After a short resting period (in the sperm-cells at least) another division phase sets in. Here each chromosome is split throughout its length, as in ordinary division, and the halves pass to daughter cells. Each cell still has the half number of chromosomes which now pass again into the nuclear resting phase. Out of these cells the mature germ-cells are produced—eggs and spermatozoa.

The generalized description just given covers the essential facts, but in

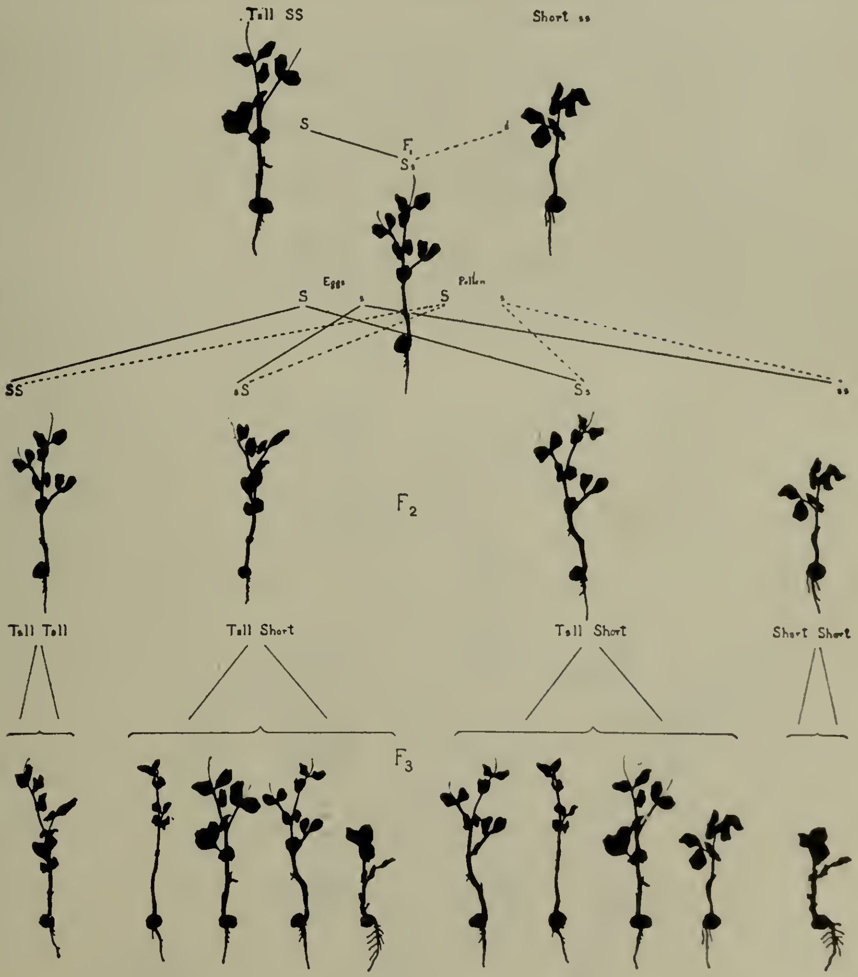


FIGURE 6

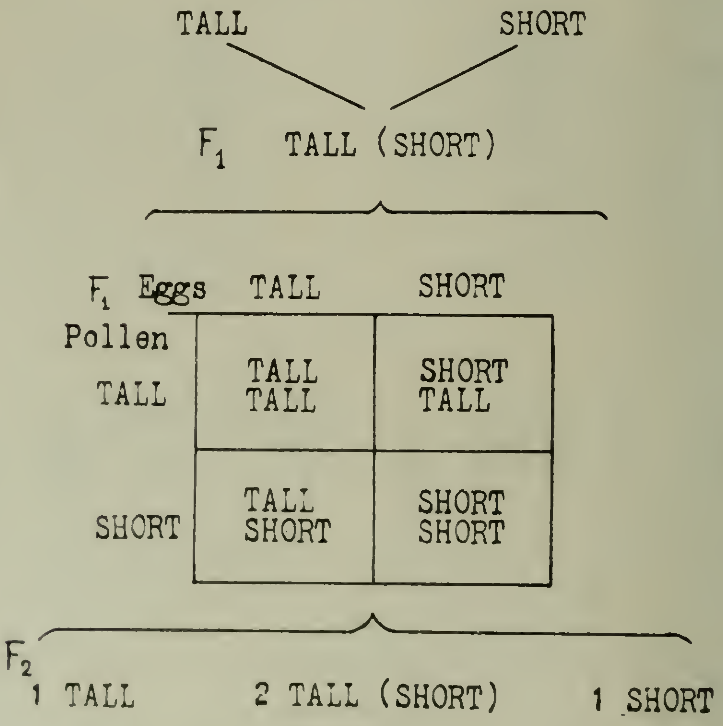
Diagram to Illustrate a Cross between Tall and Short Edible Peas

The small letter *s* stands for the short-producing gene. The large letter *S* for the normal allelomorph or tall. The offspring or "hybrid" (F_1 *Ss*) is tall. It produces two kinds of eggs and two kinds of germ-cells or gametes, viz., *S* and *s*. Self-fertilization gives three kinds of offspring (F_2) in the ratio of 1:2:1, viz.: tall tall, tall short, and short short. The first and last breed true as shown in F_3.

detail the changes differ in the ripening of eggs and sperm-cells. The sperm-cell follows closely the above scheme (Figure 3). Four spermatozoa are produced by two divisions of the original mother cell. Each sperm carries half the total number of chromosomes. The egg cell also divides twice (Figure 4), but in animals one of the products of the division is extremely small—invisible to the unaided eye. It is called a polar body. Two such bodies are "given off" by two meiotic divisions that are exactly comparable to those in the sperm except for size. The first polar body generally divides again. The three polar bodies and the egg are comparable to the four spermatozoa, but the egg alone is capable of further development. Since the first division of the chromosomes reduces the number in the egg as it does in the spermatozoa, there will be as many kinds of eggs as there are kinds of spermatozoa.

Fertilization of the eggs, each by a single sperm, leads to the union of the nucleus of the egg with that of the sperm to form a new nucleus containing the sum of the two sets of chromosomes, or the total number (Figure 5). Each divides lengthwise at every subsequent division of the segmenting egg, and, in consequence, every cell of the body comes to contain the duplex number.

The application of the chromosome mechanism to Mendel's laws may be illustrated by the following example, in which symbols for the genes will first be given (Figure 6) and then (Figure 8) chromosomes substituted



Pollen \ Eggs	TALL	SHORT
TALL	TALL TALL	SHORT TALL
SHORT	TALL SHORT	SHORT SHORT

FIGURE 7

Diagram to Illustrate the Fertilization of the Two Kinds of Eggs, Tall and Short, of the Hybrid (F_1) by the Two Kinds of Pollen Grains, Giving the Ratio 1:2:1 in F_2

for the letters. Mendel crossed a pea from a race breeding true to tallness to a pea from a short race (Figure 6). The genes for tallness may be designated by the letters *SS*. The ripe egg will contain one of them, *S*. Similarly in the other race (*ss*) each pollen grain will contain one gene for short, *s*. After crossing, the fertilized egg will contain this pair of genes, *Ss*,—one for tall and one for short. The hybrid that develops from the egg will contain the same pair in all of its cells.

The fact that the hybrid itself is tall means that, when both are present, the tall gene produces its full effect, and is said to be dominant or to dominate.

When in the hybrid the germ-cells ripen, the genes in question, tall and short, will separate, and each germ-cell will come to contain one or the other (Figure 6). The garden pea is hermaphroditic or bisexual. Eggs and pollen develop on the same plant. Half of the eggs of the hybrid will, when maturation is completed, contain the short-producing, and half will contain the long-producing, gene. The same is true for the pollen. Chance

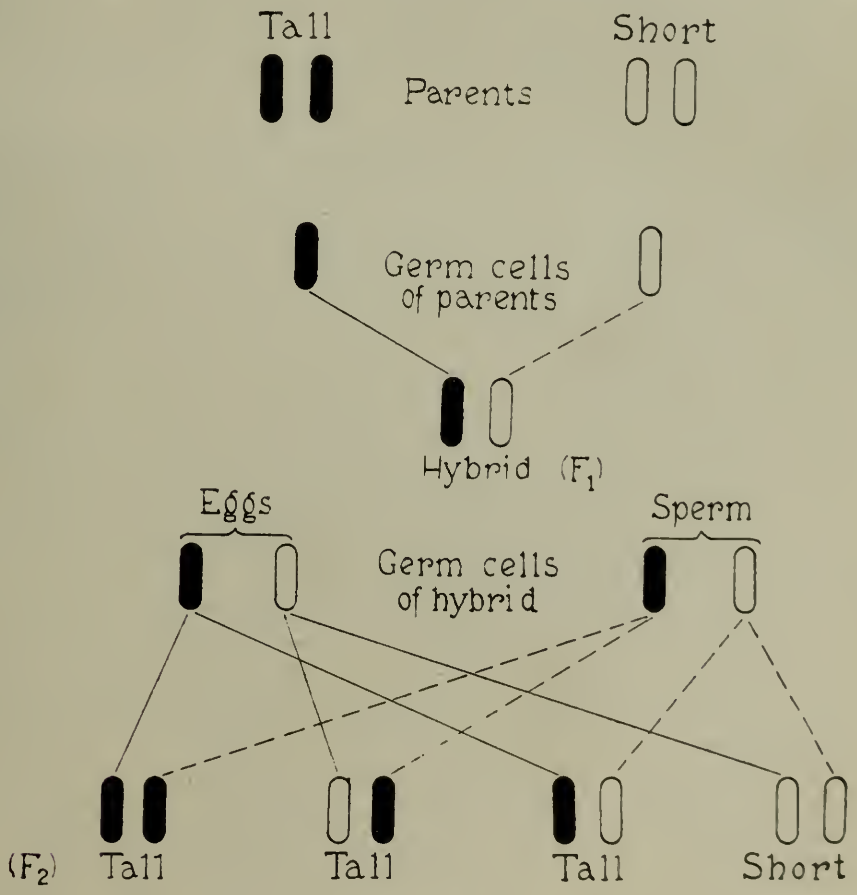


FIGURE 8

Diagram Illustrating How a Pair of Chromosomes Behaves in the Cross between a Tall and a Short Pea

The black rods are each supposed to carry the gene for tall, the white rods the one for short.

fertilization of any egg by any pollen grain will give four classes of individuals in the ratio of 1:2:1 (Figure 7).

One class will be pure for tall, since it contains both of the tall-producing genes. It will breed as true to this character as did the original tall parent, even although one of its grandparents was short.

Two classes will be hybrid. They contain one tall-producing and one short-producing gene. If self-fertilized they will behave like the hybrid parent (F_1), and give again a ratio of 1:2:1.

The remaining class will be short, since it contains the two short-producing genes. It will breed true to this character, even although its parents were tall (Figure 6).

If, instead of representing the elements (genes) for tall and short by letters, they are represented by a pair of chromosomes, the outcome (Figure 8) is the same as that just described. In this diagram the chromosomes which carry the genes for tall are represented by black rods, the homologous chromosomes, which carry the genes for short, are represented by white rods. The hybrid has one of each kind. These, separating at maturation of the germ-cells, give two kinds of reduced eggs and two kinds of reduced pollen cells. Chance fertilization of any egg by any pollen grain gives the 1:2:1 ratio.

This example shows how the chromosomes supply the mechanism for Mendel's first law, that may be called the law of segregation, which means that in the hybrid the contrasted elements are segregated into different germ-cells.

There is an adequate test of Mendel's first law by means of a back-cross. If, in the last case, the eggs of the F_1 hybrid were artificially fertilized, not by its own pollen but by pollen from a race of short peas, the expectation on the theory would be for half the offspring to be tall, half short, because, on the theory, the hybrid produces two kinds of germ-cells —tall and short—in equal numbers, and the germ-cells of the short peas are recessive and all short-producing. The offspring should then be hybrid (i.e., tall) and short in equal numbers. The results confirm the expectations. Many other tests are possible and have all been made in other Mendelian crosses. They confirm Mendel's theory.

When two independent pairs of characters enter into the same cross, the elements of each pair being carried by a different pair of chromosomes, each pair follows the first law, but, since each pair segregates independently, there will be four kinds of germ-cells that give by recombination in the second generation 16 classes. These fall into four groups in the ratio of 9:3:3:1. For example, if a pea that is tall and has purple flowers is crossed to a short pea with white flowers (Figure 9), the first generation peas are all tall and purple, since these two characters dominate. When the chromosome pairs are segregated in the hybrids, there will be four and only four assortments (germ-cells) possible, as shown in Figure 9. These are tall purple, tall white, short purple, short white, in equal numbers. If, when the hybrid is self-fertilized, the meeting of any one of the four kinds of pollen grains with any one of the four kinds of eggs is a

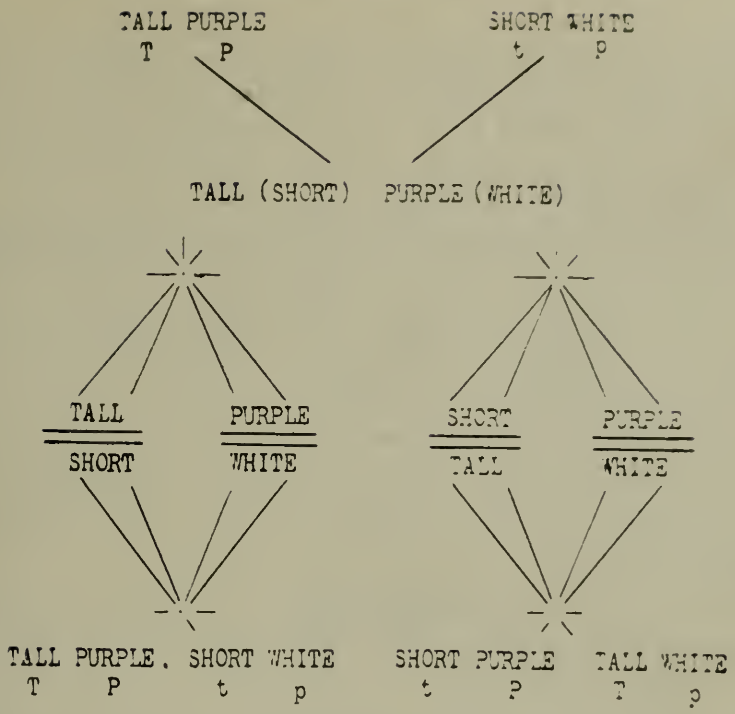


FIGURE 9

Diagram Illustrating a Cross between Tall Purple and Short White Peas

The two spindles in the lower part of the diagram show how the independent assortment of the two pairs leads to the formation of four kinds of germ-cells (*TP, tp, tP, Tp*).

		EGGS			
		Tall purple	Tall white	Dwarf purple	Dwarf white
Pollen	Tall purple	Tall purple Tall purple (9)	Tall white Tall purple (9)	Dwarf purple Tall purple (9)	Dwarf white Tall purple (9)
	Tall white	Tall purple Tall white (9)	Tall white Tall white (3a)	Dwarf purple Tall white (9)	Dwarf white Tall white (3a)
	Dwarf purple	Tall purple Dwarf purple (9)	Tall white Dwarf purple (9)	Dwarf purple Dwarf purple (3b)	Dwarf white Dwarf purple (3b)
	Dwarf white	Tall purple Dwarf white (9)	Tall white Dwarf white (3a)	Dwarf purple Dwarf white (3b)	Dwarf white Dwarf white (1)

FIGURE 10

Diagram Illustrating the 16 Classes of F_2 Individuals Resulting from a Cross between Tall Purple and Short White Peas

These 16 classes fall into four groups as determined by the presence of two or one or no dominant gene, giving the 9:3:3:1 ratio.

chance event (which means that there is no selective fertilization), the 16 possible unions result that are shown in the next diagram (Figure 10). When the characters shown by the individuals are classified according to dominance, there are present nine with two dominants, three with one, and three with the other dominant, and one with neither dominant, the so-called double recessive.

The outcome in the second generation from a cross involving two pairs of genes is the same regardless as to whether both dominants went in from one side and both recessives from the other side (as above) or whether a dominant of one pair and a recessive from the other pair go in together. Independent assortment of the pairs leads in both cases to the same kinds of germ-cells in the hybrid.

On the other hand, as will be shown later, the result would be expected to be different, as it is in fact, if two (or more) pairs of genes should be in the same chromosome pair. Mendel did not meet with cases of this kind although several are known today in garden peas and a great many such cases in other types.

When three or more independent pairs of characters are present in a cross, the principles involved are the same—there will be, of course, more classes in the second generation.

Mendel regarded one member of a pair of characters as completely dominant over the other member. Many such cases are at present known, but there are many other cases in which dominance is not complete, and the character in the hybrid lies somewhere between those of its parents. Nevertheless, by suitable tests it has been amply shown that the segregation of the genes in the hybrid is as accurate and clean-cut as in the former cases.

One of the difficulties, that has come up as our information has broadened, relates to situations where modifiers of characters are present. The simplest cases of this sort are those where specific genes exist that intensify or, conversely, dilute the character in question. In extreme cases a modifier may completely suppress the expression of another character. The occurrence of such genes is an ever present anxiety to the geneticist, but suitable methods are known by which modifiers can themselves be studied as are other Mendelian characters.

More difficult to handle are multiple-factor cases in which the development of a character is the direct result of more than one kind of gene—all, taken together, adding to or subtracting from a given measurement. Height in man or in corn are examples. The height of a man may be due to length of legs or body or neck or of all taken together or to what is known as hybrid vigor. Height in man may be due to the time at which sexual maturity is reached or to the activity of glands of internal secretion, etc. Different genetic factors may bring about these changes. Hence no one or even a few factors suffice to explain the result, and, since their effects may overlap, the analysis is made difficult. In such cases the usual resort is to employ statistical methods, but the interest of the geneticist lies in the detection of the specific genes involved. As long as a given result is the outcome of different, unknown, genetic factors, the statistical treatment may serve as a temporary expedient. When the genetic analysis can be made, it clears up the situation by resolving it into the specific agencies producing a complex result.

So far the chromosomes have been considered as wholes, but during the last fifteen years evidence has been obtained that proves that at one stage in their cycle there may occur interchanges between the members of each

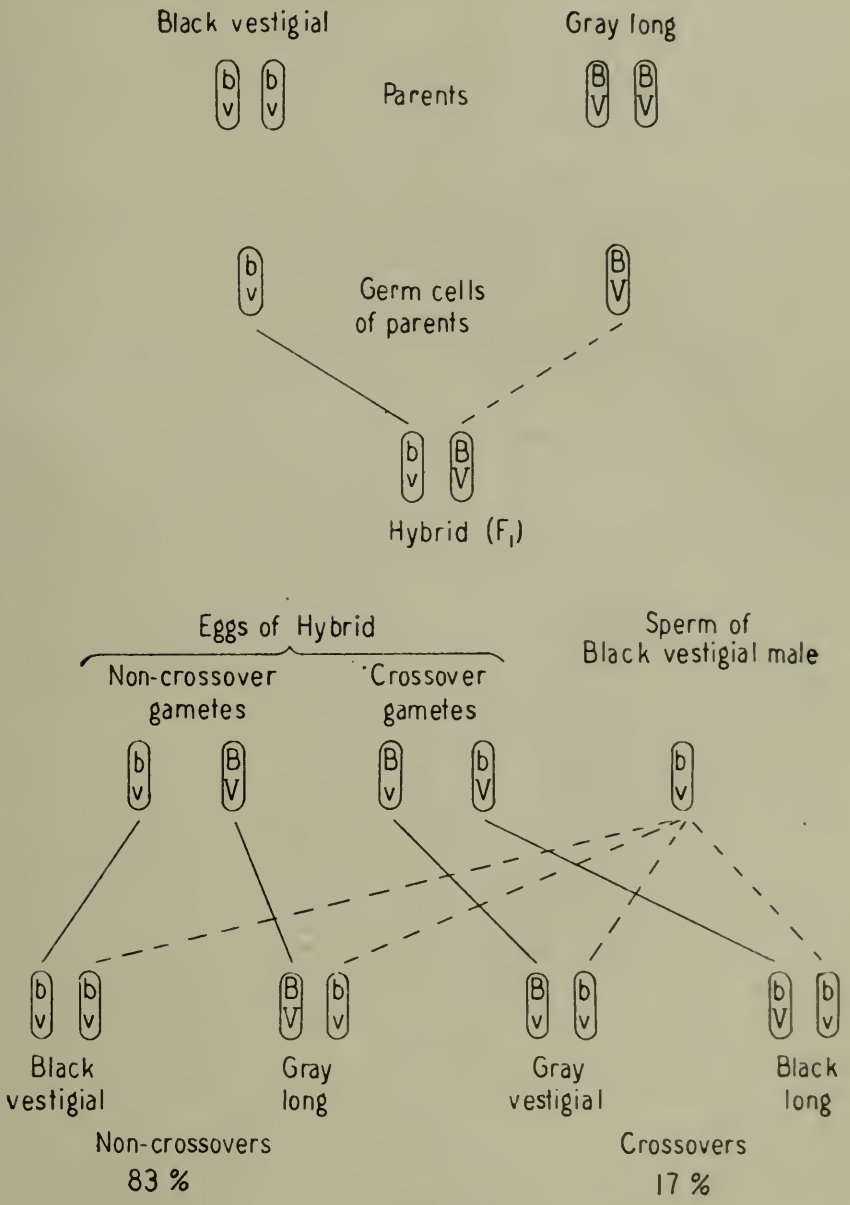


FIGURE 11

Diagram Illustrating Crossing Over between Members of a Chromosome Pair of the Hybrid from a Cross between a Black Vestigial and a Gray Long Fly

The gene for black is *b*, and for gray *B*; that for vestigial is small *v*, and that for long wings is large *V*.

pair. In so far as this takes place, Mendel's second law no longer covers the facts. The new procedure enables us, none the less, to treat the problems of heredity that arise with the same exactness as applies to Mendel's laws. This phenomenon is known as crossing over.

An example will serve to illustrate crossing over. There is a mutant race of the vinegar fly *Drosophila* that has black wings. It gives when crossed to the wild fly that has gray wings, three grays to one black in the second generation. There is another mutant race that has vestiges of wings. The wings of the wild type may be called long. Vestigial crossed to long gives three longs to one vestigial in the second generation. In other words, the two mutant characters are recessive, and when each alone is present in a cross it gives the Mendelian ratio 3:1 in F_2, and in a back-cross a 1:1 ratio.

Now, if a fly that is *both* black and vestigial is crossed to one that is gray and long, the offspring (F_1) are gray and long. If the F_1 females are mated to double recessive, black, vestigial males of stock, the next generation gives:

Gray long	Black vestigial	Gray vestigial	Black long
41.5	41.5	8.5	8.5

Since the black vestigial male used in the back-cross carried only recessive genes, the result reveals the kinds and proportions of the germ-cells of the hybrid female (F_1). The result means that 41.5% of her germ-cells contained the two recessive genes that went in from one parent and 41.5% of the germ-cells contained the two dominant genes that went in from the other parent. We speak of these genes as linked together to the extent shown by the figures. In addition there are two other classes of germ-cells that are, so to speak, interchanges between the combinations that went in together. These classes are called cross-overs and constitute 17% of all the F_2 flies.

If we suppose that black (b) and vestigal (v) are carried by the same chromosome and that gray (B) and long (V) are carried by the homologous chromosome of the other parent, then, as shown in Figure 11, the two chromosomes are segregated into different cells. If nothing further had happened, there would be only two kinds of germ-cells (bv and BV), 50% of each kind, but, as the results show, there are only about 42% of each kind. In addition there are two other kinds of egg cells that can be accounted for if a reciprocal interchange has taken place between the two chromosomes in question. This may be represented in a diagram by the two chromosomes overlapping at points between the Mendelian pairs (Figure 12). If the chromosomes fuse at the point of crossing (between the pairs) and the four ends reunite on each side of the crossing, the two resulting chromosomes will be bV and Bv respectively. It is to be noted that the interchange has involved large pieces of the chromosomes, but each will still contain its complete number of genes in the same sequence as before. As the numerical results show, the interchange takes place in only 17% of the original germ-cells.

There is another important relation that the work on crossing over has revealed. The outcome is in principle the same whether the two recessives enter from one side and the two dominants from the other, as in the case above, or whether a recessive and a dominant enter together. The classes themselves in F_2 will, of course, be different. Thus, if black and long had entered together from one individual and gray and vestigial from the other, the second generation of flies (F_2) from a back-cross gives:

Non Cross-overs		Cross-overs	
Black long	Gray vestigial	Black vestigial	Gray long
41.5	41.5	8.5	8.5

Again it is evident that the combinations that went in together stay together (linkage) more often than they interchange (cross over), and that the ratios of the different classes are the same as in the other combination. In other words, the result does not depend on the nature of the combinations that enter together but on the chromosomes as wholes. This relation has been found to hold for all cases studied.

In the example given above, the two characters involve the same organs —the wings. This means nothing in itself, for an example might equally well have been chosen in which one pair of characters involved one organ and the other a different one—a wing and an eye, for instance. In fact, the genetic work has shown that the order of the genes in the chromosomes bears no relation whatsoever to the principal effect produced by them in the individual. In this connection it is worth while to recall that since each gene may affect several different parts at the same time, and that we

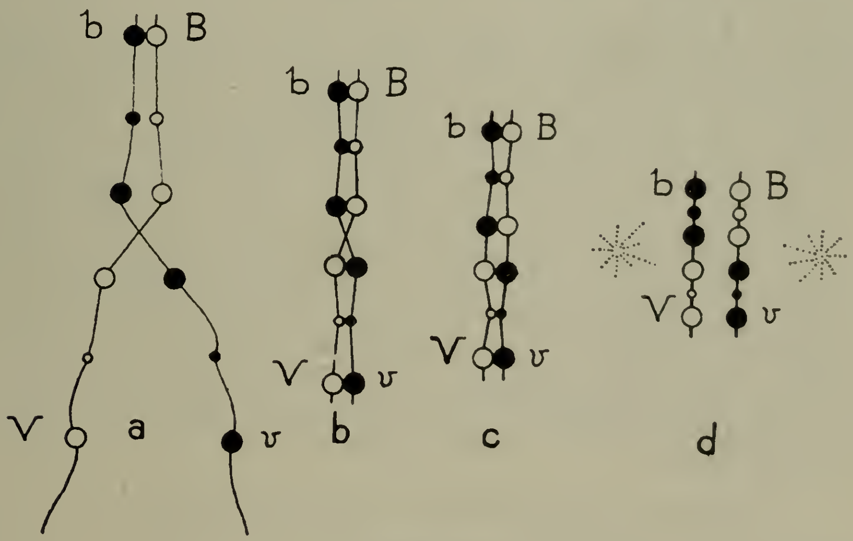


FIGURE 12

Diagram Illustrating Crossing Over between a Pair of Chromosomes, One (Black Dots) Carrying the Genes for Black (*b*) and Vestigial (*v*), the Other for Gray (*B*) and Long (*V*)

In (*a*) these two chromosomes overlap before coming together as in (*b*). In (*c*) they are represented as having come together. The two lines of genes on each side of the plane of union have now united each into a single line.

select only the major one for purposes of classification, there is no a priori reason to expect that the structure of the chromosome would or need have any corresponding relation to the structure of the individual.

The study of many pairs of genes in the same chromosome has shown that for any two pairs studied together a definite amount of crossing over takes place. Of course this varies somewhat in each case when the environments are different. Temperature may have an effect; the presence of other genes even seems in some cases to have an effect. If, as happens in extreme cases, the order of the genes becomes changed in a part of a chromosome, the effect on crossing over is marked in the region affected when one chromosome of the pair has the genes in normal sequence and its mate in the reverse. We can see, in a way, why this should happen, because the regions no longer lie opposite each other. If crossing over does take place, the resulting chromosomes will lack one or the other set of genes that are normally present, and in their absence the fly dies sooner or later.

Crossing over may be approximately the same in both sexes, but it may also be different. In extreme cases, such as *Drosophila,* there is no crossing over in the male.

It has been possible to determine the relative position of the genes in the chromosomes by taking advantage of crossing over. If we assume that, in general, crossing over is as likely to occur at one level as at any other, it follows that the nearer together two pairs of genes lie, the less the chance that crossing over will occur between them. Also, the further apart they lie, the greater this chance. If, then, we interpret the numerical results of crossing over as distance between the genes, we get a value for their position.

After getting the "distance" apart of two pairs, we proceed with a third pair and locate it in relation to the first two pairs. Proceeding in like manner, we can locate all of the genes in any linked series, i.e., in any one chromosome. In this way charts of the chromosomes have been constructed. By arbitrarily taking one end of a series as a reference point, calling it zero, numerical, serial distances or numbers can be assigned to all the other genes of that chromosome. Such a chart enables one to calculate for any new gene in a given group its relation to all the other genes in the group, after determining its crossing-over value with any other two known genes. The charts would, then, justify themselves even if the argument on which they are based were not sound, because, like Mendel's laws themselves, the charts enable us to handle genetic results on a numerical or quantitative basis.

Human Inheritance and Psychology

The modern theory of heredity touches, at many points, questions relating to human psychology, even though the inheritance of acquired behavior patterns is not one of them. Man, it is true, furnishes very poor genetic material, because the output of a given pair is too meager to furnish a characteristic sample, and, when it comes to psychological traits, so called, the diagnosis is sometimes uncertain, and a collection of materials from different matings is often too heterogeneous for accurate analysis.

In so far as structure is concerned, there can be no doubt that human characters follow Mendel's laws. There is, of course, every a priori expectation that this should be so, for in his physical characters man is like other mammals. To the extent that types of behavior are direct expressions of the gross physical characters, we might anticipate that the same laws would be found to hold, but it would be unsafe to press too far this anticipation, because man is almost unique amongst animals in his extraordinary capacity to learn. His prolonged childhood gives an opportunity to impress upon him the traditions, customs, beliefs, and prejudices of his race. Even if there were several or very many inherited backgrounds or reaction systems, it would remain to be shown to what extent these might be obliterated or expanded by training. This is a question of paramount importance for the bearing of genetics on psychology. Before attempting to discuss it, I shall bring forward some illustrations of the inheritance of typical human physical characters, since their acceptance or rejection must supply the validity for any further profitable discussion of the more complex subject of psychological traits.

The most familiar and oft-quoted case is that of brown versus blue eyes. Suppose a brown-eyed man, with an ancestry of brown eyes only, marries a blue-eyed woman. All the children have brown eyes. Two individuals that have had a similar history, mated to each other, will on the average have three brown-eyed children to one blue. One such (F_1) pair would not have enough children to make the ratio statistically valid, but when data from enough such material are collected, it is found that there is an approach to a 3:1 ratio. There are, it is true, a few apparent discrepancies that have been reported. Some of them may be due to illegitimate children or to differences in classifying the eye colors, for there are other iris colors, and their relation to brown is not understood. There are also cases where the brown is only present as a fleck. One step more and it might not be recognizable. With all due allowance for these exceptions, there is no reasonable doubt that brown and not-brown (blue) are a contrasted pair of genes.

The brachydactyl or short-fingered hand (and foot) furnishes a striking dominant character. One segment of the fingers is absent. The character breeds true to type without intergrades, and at present is known only in a hybrid condition. Whether the double dominant is lethal is not known. The individuals of the strain, having recessive or normal hands, breed true, i.e., never give brachydactyl offspring, and this is entirely consistent with the behavior of the normal hand as a Mendelian recessive in contrast with the dominant brachydactyl hand.

There are also several other structual characters in man, nearly all of which would be called malformations, that are simple cases of Mendelian inheritance, and there are many others that have not, as yet, been proved to be the effect of a single pair of genes. The medical literature contains a very large number of pedigrees showing the recurrence in successive generations of one or another bodily defect. Taken collectively they form an impressive mass of evidence showing that defective germ-plasm is constantly cropping out in man. It may seem surprising that such defects have not become more widely distributed, but in some cases the individuals may be selected against or be less productive, which would tend to eliminate the character. On the other hand, when recessive, it would be lost to sight and might be carried in the heterozygote for some time beneath the surface, and become widely spread in this way. Even then its distribution would tend to be limited. Perhaps more significant might be its proximity (linkage) to other genes that do or do not count in the viability of the individual. Until these and other possible conditions are known, it seems rather futile to give too much practical attention to them, unless by way of cautioning those who like to draw pessimistic pictures of the future of the human race. For present purposes the occurrence of such sporadic physical defects that are inheritable furnishes at least a basis for opinion as to the likelihood of similar conditions affecting the psychical make-up of human individuals.

Most races are mixtures, in different proportions, of four blood types that may be represented in Mendelian terms by *AB, Ab, aB, ab*—that is, as combinations of two pairs of genes. The data are also covered by the assumption of three allelomorphic genes, *A, B, R,* giving the four groups *AB, AA* (and *AR*), *BB* (and *BR*) and *RR*. In some respects the one formulation fits the data better, in other respects the other. In either case the evidence is sufficient to show that the differences are discontinuous and follow Mendel's principle of segregation. In passing it may be noted that, so far as known, no other subsidiary differences, either physiological or psychological, are associated with these blood types. No one of them has any known advantage over the others. The relative proportions in which they are found are probably due to racial mixtures.

The ground under our feet is less sure when we come to consider the effects of glandular disturbances either in body or mind; for, while the evidence of a close relation exists in a physiological sense, the support is very weak when we examine the case for inheritance. Cretinism is perhaps the best example. It is due to a deficiency in thyroid. The effects can be counteracted—if not too prolonged—by the administration of thyroid extract. The malady was supposed to be endemic in certain parts of the world, and is believed to be due to insufficient iodine in the food. Since in such localities only a small percentage of individuals are cretins, the question arises whether they are genetically different from the other members of the community in the sense that they are more sensitive, or less resistant, hence more easily affected by lack of iodine, or whether some environmental difference has caught a few individuals unprepared. The cretin is a physiological type accompanied by distinct psychological deficiencies. If, as stated to be the case, the mind may, after treatment, become normal, it follows either that the brain may develop to the normal stage after treatment, or that its functioning is restored with a sufficiency of thyroid secretion.

It has become popular in recent years to classify human beings, supposedly normal, according to the activity of the glands of internal secretion. One extremist calls them the "glands of personality" and tries to recognize a thyroid type, a pituitary type, and an adrenal type, which recalls the old-fashioned division of mankind into bilious and lymphatic, nervous and sanguine temperaments. Other classifiers speak of the athletic, the asthenic, and the pyknic, and ascribe to them associated peculiarities of physique and temperament. These crude efforts lend themselves to literary extravagances. Scientifically their interest rests on the more or less plausible assumption that the endocrines do play a rôle in the development of the individual, and have some part in molding both physical characters and psychological traits. There are, however, several considerations that should make us pause before accepting this evidence at its face value. First, it is probable that some of the observed results are the outcome of the combined action of several or many endocrine and hormone factors. To presuppose that the classified types are due to the preponderance of one alone of these endocrines ignores the recognized compensatory properties of these glands about which very little is known but enough to put us on our guard. Second, the effects are supposed to wax and wane in response to environmental influences that affect the glands in question. Here again we are ignorant as to how significant are the supposedly inherited differences in the functional activity of the glands, and how important are the secondary influences of the individual's personal history. Finally, practically nothing is accurately known as to the rôle of inheritance in transmitting supposedly different types. It may be freely conceded that we may expect to find by further observation and experiment that there are quantitative and qualitative differences in the glandular make-up of individuals, and that these follow the same laws of inheritance as do other bodily characters, but *at present* this anticipation furnishes a very insecure basis for the extravagances of popular writers on these topics.

There are other human types that are supposed to be inherited, but the evidence for which is unsatisfactory. There are individuals who are subnormal in regard to their capacity to learn certain technical subjects or conventions that "normal" persons can be taught. The "deficient" types are sometimes collectively called feeble-minded and sometimes classified as idiots, imbeciles, morons. There are no certain physical stigmata that characterize the imbecile or moron. Their mental development is vaguely compared with that of a child of 1-2, 3-7, 8-12 years old. It is not known whether there is only one defective type of this sort, or several, or many. If only one, and if it is recessive, it should give, when inbred, offspring like itself; but if there are different types, two imbeciles would give normal offspring in the first generation. If the condition is due to several genes, the expectation is more complex. The pedigrees are by no means consistent or convincing. There is more than a suspicion that in some cases the disorder is the result of syphilitic infection at or after birth, and possibly in other cases to infantile disease or infection, or to physical disorders of one or another kind.

The same and other possibilities apply also to the higher "grades" of defectives. Under the circumstances one may hesitate before accepting some of the far-reaching programs for eliminating these defective strains from the human germ-plasm, especially when it is realized that these, to be effective in a reasonable time, entail celibacy on all the suspected kin.

The same criticism may be made of the way in which criminality and other anti-social types have been handled as literary material for popular consumption. It is needless to point out perhaps that most of this sort of propaganda is at present outside the range of genetics even when carried on by students of genetics.

There are a few pedigrees that seem to indicate that certain types of insanity are hereditary. One type, Huntington's chorea, is probably the best authenticated case, both because it is a dominant trait and appears in each generation in individuals who have passed middle age, and also because the disease is so rare that its reappearance in successive generations in the same stock could not be due to mere chance. But even here the diagnosis is sometimes uncertain, and the same criticism applies a fortiori to several of the other forms of insanity. In fact, in some cases at least, the type of outbreak or expression shown by individuals is so different that it is by no means certain that it is traceable to what is generally spoken of as an unstable heredity or neurotic basis. The latter expression is obviously too vague to permit more than plausible guessing, which in the hands of enthusiasts is more likely to bring genetics into disrepute than it is to lead to an increase in human understanding.

It is needless to go over the many other cases of human defective types that, on the basis of rather meager pedigree charts, are claimed as cases of heredity. Taken as a whole they are interesting and convincing to the extent that hereditary factors are probably responsible for many of the defective and pathological structural characters found in the human races. In fact, all the books that deal with inheritance in man cannot fail to give the impression that nearly all of our definite information about human inheritance concerns malformations, diseases, mental defects, and trivial characters of several sorts. When we look for the other side of the picture, the heredity of beneficial traits, we usually meet with more eloquence than evidence. The difficulty here seems to be partly due to the absence of any type that may be called the normal. Without a standard for comparison the treatment becomes vague. Moreover, in human society as it exists at present it is not easy to decide that one physical type is better adapted than another, and, when it comes to deciding which emotional and intelligent types are better or worse, the situation becomes far too complex to handle with any probability of success. The socially "succesful" individual may be more a creature of his family influence and training and social opportunities than of any inherited psychical advantage over his fellows.

Under the circumstances one resort has been to apply biometrical methods to these complicated problems. Francis Galton (49-54) in his famous studies on the heredity of human traits initiated biologists into the statistical treatment of problems of heredity.[1] His books on *Hereditary Genius* and on *Natural Inheritance* are the recognized masterpieces in the literature of the subject.

In originality, in broad-mindness, in moderation of statement, Galton's work stands out above all others of his time. He recognized the limitations of his methods, and he saw far more clearly than others the complexities of the problems. His recognition of the significance of discontinuous inheritance was far in advance of his time and an important step forward in the Darwinian field—or at least would have been so had not his statements been forgotten by his followers. Galton dealt with mixed populations and attempted by statistical treatment of the materials to deduce certain laws or principles of heredity. This method of approach has advantages when it is impossible to disentangle the many factors that enter into the result, but it may be misleading, and admittedly fails to furnish the specific solution of the problem.

According to Galton: "The two parents between them contribute *on the average* one-half of each inherited faculty, each of them contributing one-quarter of it. The four grandparents contribute between them one-quarter, or each of them one-sixteenth, and so on—the sum of the series, 1/2+1/4+1/8+1/16+etc. being equal to 1." Pearson later gave slightly different values for the average contribution from each generation, and Yule formulated the law as follows: "The mean character of the offspring can be calculated with the more exactness, the more extensive our knowledge of the corresponding characters of the ancestry." When a complex character such as the height of a man supplies the raw data, the results for each generation correspond approximately with the Galton formula. Today, in the light of Johannsen's (74) work on the sizes and weights of beans, we get further insight into the relations that lead to the expression of the kind of data that are behind Galton's law. Populations are sometimes mixed, that is, made up of individuals that have different genes that affect the same character. Furthermore, the same character, size, for example, may be also influenced by environmental factors that are not inherited. In consequence, the expression of the character is only in part determined by its genes—hence its measurements are not an exact index of what the individual received from its parents or transmits to its offspring. The error may be so great as entirely to vitiate conclusions that attempt to deduce from such data the fundamental principles of heredity. In so far as the environmental factors are constant and cancel out in each generation, certain correlations between generations may be made manifest, but the exact laws of heredity that give such correlated results cannot be brought to light by this procedure. Selection in such a mixed population will in time sort out the genetically pure lines of which it is composed. The reversion toward mediocrity, one of Galton's best known contributions, finds a specific explanation in those relations that Johannsen's work brings to light.

---

[1] Quetelet had earlier utilized such methods in the measurements of human stature.

An example will serve to illustrate the difference in the treatment of a specific case when Galton's and Mendel's procedures are followed. Galton collected data on the inheritance of eye color in man and formulated conclusions as to the inheritance of brown and blue eyes. He clearly recognized the discontinuity in the contrasted eye colors and compared it with the blended type of inheritance shown by stature. Students of Mendelian heredity have also reached conclusions from the same and from similar data on eye color. The Mendelian formula states that when a brown-eyed individual that has come from a brown- and blue-eyed parent is mated to a blue-eyed individual, half the offspring will be brown- and half blue-eyed. It furnishes a hypothesis as to why this numerical ratio holds, viz.: that the brown and blue elements in the hybrid parent separate in the hybrid, and half of the germ-cells contain one element, half the other. The result does not depend on the number of brown-eyed ancestors behind one of the parents or the number of blue-eyed ancestors of the other. Here we are dealing with a specific event for which a simple hypothesis suffices. Moreover, this hypothesis, when checked up in numerous ways, gives consistent results.

If we attempt to handle the same situation by means of Galton's formula, it becomes obvious that it fails if applied to individual cases. For example, the blues extracted, in the case given above, breed true to blue. The fact that one parent had brown eyes does not enter into the result. It is true that Galton's law does not pretend to apply to single cases but states only the statistical expectation for all the offspring in a mixed population of browns and blues. Nevertheless, if it is the laws of heredity that are sought for, it is clear that Mendel's procedure is the more profitable course to pursue.

The preceding statements are not intended to imply that statistical methods of study are not of value to genetics where the theory of probability is involved, for even in Mendelian studies certain problems can be advantageously checked by recognized mathematical applications.

IDENTICAL TWINS

The most convincing evidence of the far-reaching influence of heredity in man is derived from a study of identical twins. There are, as generally recognized, two kinds of human twins, fraternal twins and identical twins. Identical twins are always of the same sex, and, as the name implies, are extremely alike in their physical traits. On the other hand, fraternal twins are no more alike physically than any other two children of the same parentage, even although born at the same time and reared under similar conditions. Fraternal twins (ordinary twins) may both be girls or both boys, or a boy and a girl.

Twin births occur about once in a hundred times. This includes both identical and fraternal twins. Certain families produce a much higher percentage of twins, and this propensity seems to be transmitted both through the father and mother to later generations. Taking all twins together, the expectation *on chance alone* for like pairs of twins and for unlike pairs of twins would be 1♀♀ : 2♀♂ : 1♂♂, or two twins of like sex to two of unlike sex. But, in fact, the actual numbers in one tabulation are 234,497♀♀ : 264,098♀♂ : 219,312♂♂, or nearly 1 : 1 : 1. This means that the occurence of identical twins is more frequent than generally supposed, since more twins of the same sex occur than chance alone calls for. It is estimated that about 1/4 to 1/5 of all twins must be identical to fulfil the requirements of these data.

The resemblance of identical twins to each other extends to the most minute details of their physical make-up. This resemblance persists throughout life, even when the persons have lived apart and under different conditions. The resemblance is, of course, generally enhanced by dressing the twins in the same way, but, while this may tend to hide any differences that they show, their similarity is entirely borne out by physical measurements.

The two most significant characteristics of identical twins, their sex and their resemblance to each other, can be accounted for by the modern theory of heredity based on certain known facts about chromosomes. In man, as in other mammals and also in several other large groups, the female has two X chromosomes and the male one X. According to a few observers the male has also a Y chromosome; according to a few others there is no Y in the male. We are familiar in other animals with situations paralleling both of these conditions. The genetic relation would be the same as long as the presence of the Y chromosome does not interfere with the effects of genes carried by the X chromosome. In *Drosophila* this holds, and, since in man the inheritance of the few sex-linked characters that are known shows no influence of the Y chromosome, we may safely assume that if present it behaves as does the Y in *Drosophila.*

This mechanism insures approximately equal numbers of males and females; since each egg, after extrusion of the polar bodies, carries one X, it will, when fertilized by an X-bearing sperm, give rise to a female (XX), and, when fertilized by a Y-bearing sperm, to a male (X or XY). The important consideration, for present purposes, is that the sex of the individual is determined at the time of fertilization of the egg. If, then, identical twins come from a single egg fertilized by one sperm—the normal rule—both individuals should always be of the same sex.

There is circumstantial evidence that in man identical twins do come from one egg. It is known in many cases that identical twins are enclosed in a common embryonic envelope, the chorionic membrane. Fraternal twins are never so enclosed, but each has its separate chorion. There is corroborative evidence of another sort. In the South American armadillo, Azara reported on hearsay, and J. Hering in 1885 confirmed the fact, that all the young of one pregnancy are of the same sex. Fernandez (45) and Newman and Patterson (119) substantiated the essential facts, the latter for the North American species, and gave further evidence concerning the origin of these armadillo embryos. The latter especially have published conclusive evidence that all the embryos of one birth develop from a single egg.

The explanation of the physical resemblance of twins rests on the following evidence: Two individuals derived from the same egg have the same chromosomes, and, since the chromosomes are the bearers of the hereditary characters, the two individuals should be alike except in so far as the environment affects these characters. The prenatal environment of man is the uterus, which for twins would seem to give as nearly homogeneous surroundings as possible, but that its effects are of secondary importance is shown by the fact that fraternal twins may at birth be physically quite different.

The environment of twins after birth may be similar, or if they are separated quite different, yet the fact that identical twins remain throughout life about as similar as at birth points to the overwhelming influence of the inheritance of physical traits.

Why two embryos occasionally develop from a single fertile egg is not understood, even although by artificial treatment twinning can sometimes be brought about in other vertebrates. Several years ago it was supposed that an explanation had been found. In certain animals, sea-urchins, Amphioxus, salamanders, etc., two embryos can be produced from one egg by separating the two cells of the two-cell stage. The eggs of mammals divide into two parts. This stage has not been seen for the human ovum, but there can be no doubt that it follows the same course as do the eggs of other mammals.

Despite this experimental evidence it seems more probable that twinning in man may come about in the same way as it does in the armadillo, where it has been definitely shown by Newman and Patterson that the doubling first appears later than the two-cell stage. In the armadillo the primordium of the four embryos is laid down at four equidistant points of the young blastoderm. The cause of this duplication is, as I have said, not known. It seems improbable that it is due to the partial separation of the first four cells, although this has not been disproved. Therefore, if a property of this sort, whatever its origin, exists in one mammal, it seems better, for the present at least, to refer the occasional human twinning to the same source.

In other vertebrates also, in fish, reptiles, and birds, more than one embryo may begin to develop on a single egg. These are often united in various combinations—two heads, two heads and bodies, etc. In man, too, similar united embryos have been frequently reported. These are in all cases identical twins—diplopagi. Their occurrence is a further argument for the view that separated identical twins come from a single egg.

The chromosome mechanism also helps us to understand why fraternal twins, and children in general, are different from one another, even when they have the same parents. Without going too much into detail, I may state that the explanation rests in part on the way in which the chromosomes are redistributed after they have paired at the ripening stage of which I have spoken. The members of each pair are distributed at random to the daughter cells at the last or maturation division of the germ-cells. Half of the chromosomes have come from one parent and half from the other parent. We suppose those from one parent carry different units from the parent, and we can calculate how many possible combinations of chromosomes are possible.

Let us suppose that a species has only two chromosomes, one pair. Each mature egg and sperm gets one or the other. If one of them is different from the other in some hereditary unit, call one chromosome A, the other a, then each ripe germ-cell (each egg and each sperm) will contain one or the other, and when they are recombined again at fertilization there will be three kinds of recombinations, or $AA+Aa+aa$, in the ratio of 1 to 2 to 1.

In a species with *two pairs* or four chromosomes (each pair having one or more differences), there will be nine classes.

Man has 48 chromosomes, which gives 282,429,536,481 possible kinds of individuals on the assumption that there are differences in the units of each pair. This is not a very improbable assumption since man is a most extraordinary mixture of many kinds of characters.

The chances, then, of any two individuals' (coming from different eggs and sperms) being identical would be about one in a billion. Of course, there would actually be many individuals more or less alike for a large number of hereditary units, and this fits in with what we see in our fellow-men—some of them do look alike even when they have a different parentage. Moreover, in general, we should expect *on the average* that members of the same family would more often resemble each other than they do other groups, and this too is true.

If, then, as a result of free assortment, the number of different kinds of human beings is very large, the number may be very much larger if crossing over takes place between members of the same pair of chromosomes. Crossing over is known to occur in other animals and in plants. In the female *Drosophila* it occurs in about 57% of the gametes for the sex chromosome, and considerably more for the two other large pairs. In man there are no certain cases of interchanges of this kind, but this probably means that as yet there are no recorded cases where two pairs of characters are known to be in the same linkage group. Our lack of information is due partly to the relatively small number of cases in man where the inheritance is definitely known, and partly to the large number of chromosomes resulting in the dispersion of human characters amongst them. Furthermore, the number of offspring from a pair is generally too small to give sufficient information and, as stated before, the diagnosis is often too uncertain when more than single matings are brought together.

The question is often asked why, if unit characters are discontinuous in inheritance, we do not easily recognize such characters in man, the details of whose physiognomy and bodily structure serve as the basis for our ability to distinguish individuals from each other. There are many reasons why we cannot often pick out such characters. A number of mutant characters are known to overlap the "normal" and may grade by invisible degrees into each other in their expression, even although the genes segregate as cleanly as in more typical cases. The hybrid form or heterozygous character is in many cases intermediate between the dominant and recessive character and may overlap one or both. Specific modifiers may increase the difficulty. It frequently happens that many factors affect the same organ, and until they can be isolated and studied apart their presence obscures the results. In other animals and plants they can be isolated by suitable matings, but in man this has not been feasible. Again, the same genes may occur in several forms (allelomorphs), each one producing a slightly different effect and any two of them combined an intermediate effect. Until they are isolated by suitable tests, the identification of the corresponding characters is difficult. It has been suggested that the difference in the brown hair of man is such a case. It requires, then, in most cases of mixed material, more information than we now possess for human traits to discover the unit characters. This kind of work can better be done on forms whose breeding is under experimental control.

The Rôle of Nature and Nurture in Human Behavior

Confronted with the evidence from identical twins showing that heredity controls the physical characters of man down to the most minute details, the student of psychological human traits may ask to what extent does this conclusion apply to his problems. We face the fact that man is almost unique amongst animals in his extraordinary ability to learn. His prolonged childhood gives an opportunity to impress upon him the particular associations to which he has been subjected. Are these extraneous influences powerful enough to submerge whatever differences may be inherited by each individual? The difficulties of reaching a decision are at present almost insuperable. There are students of heredity—extremists—who insist that all differences in behavior are due to inheritance, i.e., to the kind of reaction system that the genetic constitution supplies. There are others who are equally positive that our behavior is due entirely, or almost so, to our upbringing—the environment of the child, i.e., his early experiences, his training, education, or his imitative capacities. Most psychologists, I suppose, and certainly most geneticists, who have given serious thought to this situation, will stand somewhere between these antagonists.

The older procedure of assigning innate individual psychological differences to constitutional backgrounds, in the sense of being inborn, has been seriously questioned. The inspirational source of such a point of view does not recommend itself as a scientific method of getting reliable data. The wide differences of opinion, for example, as to how many "instincts" are common to all human beings—not to mention individual variants of these—has not led to a satisfactory outcome and has raised the question as to whether this is not clever dialectic rather than modern science.

The position taken by extremists on the other side, of whom Watson is perhaps the outstanding example, has also raised doubts and questions in the minds of the students of heredity. Speaking for myself, the experimental approach that Watson insists upon makes the strongest appeal as the procedure that will best serve to supply common ground for psychology and genetics. This method is at present being used extensively with plants and animals other than man, but for the moment I have in mind rather the work on human behavior. The question here is to what extent the evidence relating to physical individual differences supports or does not support the contention that, to the same degree, individual differences in human behavior have the same innate or inherited or congenital background. The problem is difficult for two reasons: first, because crucial evidence is almost entirely lacking; and, second, because there is sufficient evidence, I think, to show that many traits of human conduct are only acquired through association. The point at issue is, then, even if constitutional differences be admitted, whether and how far they may be completely modified by the establishment of associations, conditioned reflexes, etc., in early life.

The discussion of the problem is full of pitfalls and there is the widest possible divergence of predilections and prejudices that make an independent attitude well-nigh impossible. One approach is through a consideration of the influence of those hereditary physical differences that seem to have a significant influence on the behavior of the individual. For example, the effects of internal secretions are generally conceded to be a case in point. As already stated, their excessive presence or absence at any time may be a factor in the kind of reactions shown by the individual that may become thereby reinforced and later habitual. If there are inherited differences in the glands that produce these hormones, it may be argued that there are corresponding psychological patterns of reaction. But until it can be shown—leaving aside pathological cases—that there are such differences, and until it can be shown that the types of behavior of normal individuals that have been referred to these differences are not really due to earlier associations, the argument hangs in the air.

There is another related question that should not be forgotten. Most individuals meet in the course of their lives a vast number of situations to which they react. It is, in my opinion, premature to assume that the associations formed in babyhood may not be changed or completely reversed later, and, if so, an entirely new set of reactions established for the time being. If this can take place, it does not seem unlikely that the type of reaction existing at any one time will often be the one that has been along the line of least resistance, which might be assumed to be the one to which the physical type most easily adjusts itself.

How far the general physique (tall or broad, fat or thin, etc.) of the individual is an important factor in determining his set of psychological reactions is presumably a question for psychologists to discuss. From the genetic point of view there can be little doubt but that most of these bodily differences are inherited. Granted that physique furnishes an element of behavior, it remains to be proved that in the long run the adjustments of the individual will be along the line of reaction correlated with physique, rather than with very early associations or with predominating social influences in later life. Everyone is familiar with men of very different physique living outwardly the same kinds of lives in general. Even if social restraint is admitted as the most outstanding explanation of the result, it does not follow that there may still have been inherited differences that have been submerged. It is also conceivable in a race predominatingly of a given physical type that the social system followed may in time come to have an intimate relation to the physical type in certain economic or climatic environments. All these possibilities have been at one time or another exploited by writers of all shades of opinion and have furnished the would-be well-wishers of science with material for popular exposition.

The occurrence of identical twins has supplied the biologists with materials that convincingly establish the far-reaching effects of physical inheritance in man. Genetics has made clear why the two identical twins differ from fraternal twins and sibs in general. It follows that identical twins, if studied experimentally, should be expected to supply excellent material to test the significance in behavior of hereditary psychic factors. Galton was amongst the first to take advantage of this opportunity. He records in his *Inquiries into Human Faculty* (49) that through a questionnaire he had about 80 returns of cases of close similarity. He gives many interesting anecdotes about the similarities in the traits of some of the identical twins, but, since in most cases the twins were for several years reared together, this kind of evidence is inconclusive. Nevertheless, taken as a whole, it cannot fail to leave the impression that the similarities in some at least of the psychic differences are not due to the environment, for, when comparison with fraternal twins is made, no such close resemblance is found. Galton's comment on his evidence is interesting: "The last point to which I shall allude regards the tastes and dispositions of the thirty-five pairs of twins. In sixteen cases—that is, in nearly one-half of them—these were described as closely similar; in the remaining nineteen they were much alike, but subject to certain named differences. These differences belonged almost wholly to such groups of qualities as these: The one was the more vigorous, fearless, energetic; the other was gentle, clinging and timid; or the one was more ardent, the other more calm and placid; or again, the one was the more independent, original and self-contained; the other the more generous, hasty and vivacious. In short, the difference was that of intensity or energy in one or other of its protean forms; it did not extend more deeply into the structure of the characters. The more vivacious might be subdued by ill-health, until he assumed the character of the other; or the latter might be raised by excellent health to that of the former. The difference was in the keynote, not in the melody."

Galton also comments on the late histories of the twins that had been "closely similar" up to their early manhood and womanhood. In some cases the resemblance in "body and mind" had continued unaltered up to old age notwithstanding very different conditions of life. In other cases the parents ascribed such discontinuity as was found to some form of illness. "In not a single instance have I met with a word about the growing dissimilarity being due to the action of the firm free will of one or both of the twins, which had triumphed over natural tendencies; and yet a large proportion of my correspondents happen to be clergymen, whose bent of mind is opposed, as I feel assured from the tone of their letters, to a necessitarian view of life." His general summing up is: "We may, therefore, broadly conclude that the only circumstance, within the range of those by which persons of similar conditions of life are affected, that is capable of producing a marked effect on the character of adults, is illness or some accident which causes physical infirmity. The twins who closely resembled each other in childhood and early youth, and were reared under not very dissimilar conditions, either grow unlike through the development of natural characteristics which had lain dormant at first, or else they continue their lives, keeping time like two watches, hardly to be thrown out of accord except by some physical jar. Nature is far stronger than Nurture within the limited range that I have been careful to assign to the latter."

A recent and more detailed study of the psychological traits of a pair of identical twins has been made by Muller (108). The babies had been separated when two weeks old, and did not see each other until they were 18, and afterwards only occasionally. They were 30 years old when the tests were made. Two intelligence tests were used, the Army Alpha Test and the Otis Advanced Intelligence Test. The total scores were almost identical, indicating superior intelligence. On the other hand, when other methods were used that are supposed to test "volitional traits," the results were different, and show, Muller thinks, the influence of the training or experience that each twin had undergone. Muller concludes that most sections of the intelligence tests show a high correlation with the genetic basis of intelligence so far established "when they are applied to individuals brought up in the same general territory and social class," which is the case with these twins. If there were no such genetic basis, the chance for these twins to have been so much alike would be less than 1 in 200. In regard to the non-intelligence tests, the evidence is in the opposite direction and may indicate differences in past experience, or the tests may be inadequate for the purpose of finding genetic similarities. Nevertheless, it is obvious that identical twins should be expected to furnish the most available materials for a scientific examination of the rôle of nature versus nurture in human psychology. Instead, however, of leaving the situation to the chance discovery of identical twins reared under different environments, a direct attack on the problem is called for where the individuals can be placed, as soon after birth as possible, in entirely different social environments under expert control.

CONCLUSIONS

If the evidence discussed in the preceding review makes it seem probable that the doctrine of the inheritance of acquired characters gives no secure basis on which to bring about an intimate alliance between zoölogy and psychology, and if the statistical method developed by Galton and the

Pearson school of biometricians, when applied to mixed populations, fails to give us the specific information that Mendel's principles supply, genetics may still have indirect bearings on the study of human psychology.

The psychological problems are immeasurably complicated by the fact that human behavior is influenced to a great extent by associations, conditioned reflexes, and experiences in general that the social environment of man effects from babyhood through all the formative periods of man's life. Until more definite information is obtainable concerning mental traits and emotional reactions, some of us remain extremely skeptical of the crude and often forced attempts that have been made so far to determine what is inherited and what is acquired after birth.

A comparison of the facts of Mendelian heredity relating to plants and animals with what is known of human heredity leaves no doubt, as far as physical characters are involved, that the same rules apply to both. The extraordinary physical resemblance of identical twins to each other establishes on the soundest basis the fact that heredity plays an all-important rôle in the inheritance of structural characters. But it is extremely hazardous to carry over this inference to the psychic character of man where there is no certainty as to what extent his behavior is determined by heredity and by environment. Until suitable methods have been applied that are discriminative, conclusions will be largely a question of personal opinion or conviction. It need scarcely be urged again that identical twins under controlled conditions may supply much of the needed information.

REFERENCES

1. Adami, J. G. An address on the theories of inheritance, with special reference to the inheritance of acquired conditions in man. *Brit. Med. J.*, 1901, No. 2109.
2. Ammon, O. Die Vererbung erworbener Eigenschaften. *Naturwiss. Woch.*, 1895, **10**.
3. Assheton, R. An account of a blastodermic vesicle of the sheep of the seventh day, with twin germinal areas. *J. Anat. & Physiol.*, 1898, **32**.
4. Bateson, W. Mendel's principles of heredity, a defence. Cambridge, 1902.
5. ————. An address on Mendelian heredity and its application to man. *Brit. Med. J.*, 1906, **11**.
6. Bernstein, F. Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Zsch. Abst.-Vererb.*, 1925, **37**.
7. Brauer, J. Die Konstitutionelle Disposition zu inneren Krankheiten. Berlin, 1924.
8. Brewer, W. H. On the hereditary transmission of acquired characters. *Agric. Sci.*, 1892, **6**; 1893, **7**.
9. Brown-Séquard, —. Nouvelles recherches sur l'épilepsie due à certaines lésions de la moelle épinière et des nerfs rachidiens. *Arch. de Physiol. norm. & path.*, 1869, **2**.
10. ————. Faits nouveaux concernant la physiologie de l'épilepsie. *Arch. de Physiol. norm & path.*, 1870-1871, **3**.
11. ————. Remarque sur l'épilepsie causée par la section du nerf sciatique chez les cobayes. *Arch. de Physiol. norm. & path.*, 1870-1871, **3**.

12. ———————. Quelques faits nouveaux relatifs à l'épilepsie qu'on observe à la suite de diverses lésions du système nerveux chez les cobayes. *Arch. de Physiol. norm. & path.*, 1872, **4**.

13. ———————. Transmission par hérédité de certaines altérations des yeux chez les cobayes. *Gaz. Med. de Paris*, 1880.

14. ———————. Faits nouveaux établissant l'extrême fréquence de la transmission, par hérédité, d'états organiques morbides produits accidentellement chez les ascendants. *C. r. Acad. Sci.*, 1882, **94**.

15. ———————. Hérédité d'une affection due à une cause accidentelle. Faits et arguments contre les explications et les critiques de Weismann. *Arch. de Physiol.*, 1892, **24**.

16. ———————. Transmission héréditaire de caractères acquis. *Arch. de Physiol.*, 1893, **25**.

17. Butler, S. Life and habit. London, 1878.

18. ———————. Evolution, old and new. (2nd ed.) London, 1882.

19. Champy, Ch. Les caractères sexuels considérés comme phénomènes du développement et dans leurs rapports avec l'hormone sexuelle. Paris, 1924.

20. Cope, E. D. The primary factors of organic evolution. Chicago, 1896.

21. Crew, F. E. A. Organic inheritance in man. London, 1927.

22. Cuénot, L. L'ovaire du tatou et l'origine des jumeaux. *C. r. Soc. biol.*, 1903, **55**.

23. Cunningham, J. T. The evolution of flat-fishes. *Nat. Sci.*, 1892, **1**.

24. ———————. The origin of species among flat-fishes. *Nat. Sci.*, 1895, **6**.

25. ———————. Lyell and Lamarckism. *Nat. Sci.*, 1896, **8**.

26. ———————. Sexual dimorphism in the animal kingdom. London, 1900.

27. Danforth, C. H. Resemblance and difference in twins. *J. Hered.*, 1919, **10**.

28. Davenport, C. B. A strain producing multiple births. *J. Hered.*, 1919, **10**.

29. ———————. Influence of the male in the production of human twins. *Amer. Natur.*, 1920, **54**.

30. ———————. Heredity of twin births. *Proc. Soc. Exper. Biol. & Med.*, 1920, **17**.

31. ———————. Influence of endocrines on heredity. *Proc. & Addr. Amer. Asso. Stud. Feeble-Mind.*, 1924, Session 48.

32. Delage, M. Y. L'hérédité et les grands problèmes de la biologie générale. (2nd ed.) Paris, 1903.

33. Detlefsen, J. A. Are the effects of long-continued rotation in rats inherited? *Proc. Amer. Phil. Soc.*, 1923, **62**.

34. ———————. The inheritance of acquired characters. *Physiol. Revs.*, 1925, **5**.

35. Deutschmann, —. Ueber Vererbung erworbener Augenaffectionen bei Kaninchen. *Klin. Monatsbl. f. Augenhk.*, 1880, **18**.

36. Dingfelder, J. Beitrag zur Vererbung erworbener Eigenschaften. *Biol. Zentbl.*, 1887, **7**; 1889, **8**.

37. Dupuy, E. De la transmission héréditaire des lésions acquises. *Bull. biol. Fr. et Belg.*, 1890, **22**.

38. Eimer, G. H. Th. Organic evolution. London, 1890.

39. Eldridge, S. The organization of life. New York, 1925.

40. Emery, C. Gedanken zur Descendenz und Vererbungstheorie: IX. Variationsrichtungen und Germinalselektion. *Biol. Zentbl.*, 1897, **17**.

41. Ewart, J. C. The Penycuik experiments. London, 1899.

42. ————. Experimental contributions to the theory of heredity. Reversion and telegony. *Trans. Highland & Agric. Soc. Scotland,* 1901. Pp. 54.

43. Fairchild, D. Twins. *J. Hered.,* 1919, **10**.

44. Farabee, W. C. Inheritance of digital malformations in man. *Memoirs of Peabody Museum, Yale,* 1905, **3**.

45. Fernandez, M. Beiträge zur Embryologie der Gürteltiere. I. Zur Keimblätterinversion und spezifischen Polyembryonie der Mulita (*Tatusia hybrida* Desm.) *Morph. Jahrb.,* 1909, **39**.

45*a*. Findlay, G. F. The effect of different species' lens antisera on pregnant mice and rats and their progeny. *Brit. J. Exper. Biol.,* 1924, **1**.

46. Fischer, E. Experimentelle Untersuchungen über die Vererbung erworbener Eigenschaften. *Allg. Zsch. f. Entom.,* 1901, **6**.

47. Forel, A. Richard Semon's Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens. *Arch. f. Rassen- u. Gesellschbiol.,* 1905, **2**.

48. Frantz, M. H. Dementia praecox in twins. *J. Nerv. & Ment. Dis.,* 1919, **50**.

49. Galton, F. Inquiries into human faculty and its development. London, 1883.

50. ————. Natural inheritance. London, 1889.

51. ————. Hereditary genius. London, 1892.

52. ————. The average contribution of each several ancestor to the total heritage of the offspring. *Proc. Roy. Soc. London,* 1897, **61**.

53. ————. The possible improvement of the human breed under the existing conditions of law and sentiment. *Nature,* 1901, **64**.

54. ————. Eugenics: its definition, scope and aims. *Papers of Sociol. Soc.* London, 1905, **1**; 1906, **2**.

55. Gates, R. R. Heredity and eugenics. London, 1923.

56. Giard, A. L'hérédité des modifications somatiques. *Rev. scient.,* 1890, **46**.

57. ————. Controverses transformistes. Paris, 1904.

58. Gowen, J. W. Identical twins in cattle. *Biol. Bull.,* 1922, **42**.

59. Griffith, C. R. Are permanent disturbances of equilibration inherited? *Science,* 1922, **56**.

60. Guyer, M. F. Being well-born. Indianapolis, 1916.

61. ————. Immune sera and certain biological problems. *Amer. Natur.,* 1921, **55**.

62. ————. Orthogenesis and serological phenomena. *Amer. Natur.,* 1922, **56**.

63. ————. The germ cell and serological influences. *Proc. Amer. Phil. Soc.,* 1923, **62**.

64. Guyer, M. F., & Smith, E. A. Studies on cytolysins. I. Some prenatal effects of lens antibodies. *J. Exper. Zoöl.,* 1918, **26**.

65. ————. Studies on cytolysins. II. Transmission of induced eye defects. *J. Exper. Zoöl.,* 1920, **31**.

66. Hering, E. Ueber das Gedächtniss als eine allgemeine Function der organisirten Materie. Vienna, 1870.

67. Hertwig, O. Ergebnisse und Probleme des Zeugungs- und Vererbungslehre. Jena, 1905.

68. Hertwig, O., & Hertwig, G. Allgemeine Biologie. (6th & 7th ed.) Jena, 1923.

69. Huxley, J. S., & Carr-Saunders, A. M. Absence of prenatal effects of lens-antibodies in rabbits. *Brit. J. Exper. Biol.,* 1924, **1**.

70. Hyatt, A. The phylogeny of an acquired characteristic. *Proc. Amer. Phil. Soc.*, 1894, **32**.

71. Jenkins, R. L. Twin and triple birth ratios. *J. Hered.*, 1927, **18**.

72. Jenkinson, J. W. A reinvestigation of the early stages of the development of the mouse. *Quar. J. Micr. Sci.*, 1903, **43**.

73. Jhering, H. von. Ueber die Fortpflanzung Gürteltiere. *Sitz. preuss. Akad. Wiss.*, 1885, **105**.

74. Johannsen, W. Ueber Erblichkeit in Populationen und in reinen Linien. Jena, 1903.

75. Jones, D. F. Genetics in plant and animal improvement. New York, 1925.

76. Kammerer, P. Die Abstammungslehre. Jena, 1911.

77. ————. The inheritance of acquired characteristics. New York, 1924.

78. Kidd, W. Use-inheritance—illustrated by the direction of hair on the bodies of animals. London, 1892.

79. ————. The sense of touch in mammals and birds. London, 1907.

80. Kraepelin, E. Dementia praecox and paraphrenia. Edinburgh, 1919.

81. Le Dantec, F. L'hérédité, clef des phénomènes biologiques. *Rev. gén. scient.*, 1900, **11**.

82. ————. Les influences ancestrales. Paris, 1906.

83. Lendl, A. Hypothese über die Entstehung von Soma- und Propagationszellen. Berlin, 1889.

84. Lenz, F. Ueber die krankhaften Erbanlagen des Mannes und die Bestimmung des Geschlechts beim Menschen. Jena, 1912.

85. Leslie, G. On hereditary transmission of disease. *Edinburgh Med. J.*, 1882, (March).

86. Lillie, F. R. Supplementary notes on twins in cattle. *Biol. Bull.*, 1923, **44**.

87. Little, C. C., & Bagg, H. J. The occurrence of four inheritable morphological variations in mice and their possible relation to treatment with X-rays. *J. Exper. Zoöl.*, 1924, **41**.

88. Lush, J. L. Twinning in Brahma cattle. *J. Hered.*, 1924, **15**.

89. MacAuliffe, L. Les mechanismes intimes de la vie. Paris, 1925.

90. MacDowell, E. C. The influence of parental alcoholism upon habit formation in albino rats. *Proc. Soc. Exper. Biol. & Med.*, 1919, **16**.

91. MacDowell, E. C., & Vicari, E. M. Alcoholism and the behavior of white rats. I. The influence of alcoholic grandparents upon maze-behavior. *J. Exper. Zoöl.*, 1921, **33**.

92. ————. Alcoholism and the behavior of white rats. II. The maze-behavior of treated rats and their offspring. *J. Exper. Zoöl.*, 1923, **37**.

93. Maciesza, A., & Wrzosek, A. Experimentelle Untersuchungen über die Vererbung der durch Ischiadicusverletzung hevorgerufenen Brown-Séquardschen Meerschweinchen-Epilepsie. *Arch. f. Rassen- u. Gesellschbiol.*, 1911, **8**.

94. McCordock, H. A., & Congdon, C. C. Suppurative otitis of the albino rat. *Proc. Soc. Exper. Biol. & Med.*, 1924, **22**.

95. McDougall, W. An experiment for the testing of the hypothesis of Lamarck. *Brit. J. Psychol.*, 1927, **17**.

96. Mendel, G. J. Versuche über Pflanzen Hybriden. *Verh. Naturf. Verein in Brünn*, 1865, **4**.

97. Meyer, S. Gedächtnis und Vererbung. *Arch. f. Rassen- u. Gesellschbiol.*, 1906, **3**.

98. Miles, M. Heredity of acquired characters. *Amer. Natur.*, 1892, **26**.

99. Mohr, O. L. Mikroskopische Untersuchungen zu Experimenten über den Einfluss der Radiumstrahlen und der Kältewirkung auf die Chromatinreifung und das Heterochromosom bei *Decticus verruccivorus* (♂). *Arch. f. mikr. Anat.*, 1919, **92**.

100. Morgan, T. H. The modern theory of genetics and the problem of embryonic development. *Physiol. Revs.*, 1923, **3**.

101. ————. Mendelian heredity in relation to cytology. General cytology, Sec. XI. Chicago, 1924.

102. ————. Are acquired characters inherited? *Yale Rev.*, 1924, (July).

103. ————. Human inheritance. *Amer. Natur.*, 1924, **58**.

104. ————. Evolution and genetics. Princeton, 1925.

105. ————. Genetics and the physiology of development. *Amer. Natur.*, 1926, **60**.

106. ————. The theory of the gene. New Haven, 1928.

107. Morgan, T. H., Sturtevant, A. H., Muller, H. J., & Bridges, C. B. The mechanism of Mendelian heredity. New York, 1923.

108. Muller, H. J. Mental traits and heredity. *J. Hered.*, 1925, **16**.

109. Murray, G. R. A study of twins in health and disease. *Lancet*, 1925, **208**.

110. Myerson, A. The inheritance of mental diseases. Baltimore, 1925.

111. Newman, H. H. The ovum of the nine-banded armadillo. Growth of the ovocytes, maturation and fertilization. *Biol. Bull.*, 1912, **23**.

112. ————. The modes of inheritance of aggregates of meristic (integral) variates in the polyembryonic offspring of the nine-banded armadillo. *J. Exper. Zoöl.*, 1913, **15**.

113. ————. The natural history of the nine-banded armadillo of Texas. *Amer. Natur.*, 1913, **47**.

114. ————. Heredity and organic symmetry in armadillo quadruplets. I. Modes of inheritance of band anomalies. *Biol. Bull.*, 1915, **29**.

115. ————. Heredity and organic symmetry in armadillo quadruplets. II. Mode of inheritance of double scutes and a discussion of organic symmetry. *Biol. Bull.*, 1916, **30**.

116. ————. The biology of twins. Chicago, 1917.

117. ————. The physiology of twinning. Chicago, 1923.

118. ————. Evolution, genetics and eugenics. Chicago, 1927.

119. Newman, H. H., & Patterson, J. T. A case of normal identical quadruplets in the nine-banded armadillo, and its bearing on the problems of identical twins and of sex determination. *Biol. Bull.*, 1909, **17**.

120. ————. Development of the nine-banded armadillo from the primitive streak stage to birth: with special reference to the question of specific polyembryony. *J. Morph.*, 1910, **21**.

121. ————. The limits of hereditary control in armadillo quadruplets: a study of blastogenic variation. *J. Morph.*, 1911, **22**.

122. Ogilvie, G. Some remarks on the inheritance of acquired immunity. *Brit. Med. J.*, 1901, No. 2105.

123. Ornstein, B. Ein Beitrag zur Vererbungsfrage individuel erworbener Eigenschaften. *C. B. Deutsch. Ges. Anthrop. Ethnog. Munich,* 1889, **20**.

124. Orr, H. B. A theory of development and heredity. London, 1893.

125. Orth, J. Ueber die Entstehung und Vererbung individueller Eigenschaften. Festschrift Kölliker. 1887.

126. Osborn, H. F. Palaeontological evidence for the transmission of acquired characters. *Rep. Brit. Asso.*, 1889; *Nature*, **41**; *Amer. Natur.*, 1889, **23**.

127. ————. Are acquired variations inherited? *Amer. Natur.*, 1891, **25**.

128. ————. Present problems in evolution and heredity. *Med. Rec.*, 1892, p. 71.

129. ————. The hereditary mechanism and the search for the unknown factors of evolution. *Biol. Lectures, Woods Hole*, 1895, **3**.

130. ————. Ontogenetic and phylogenetic variation. *Science*, 1896, **4**.

131. PACKARD, A. S. The life history of certain moths of the family Cochliopodidae with notes on their spines and tubercles. *Proc. Amer. Phil. Soc.*, 1893, **31**.

132. ————. On the inheritance of acquired characters in animals with a complete metamorphosis. *Proc. Amer. Acad. Sci.*, 1894, **29**.

133. PARKER, G. H. Identical twins with dementia praecox. *J. Hered.*, 1926, **17**.

134. PAULY, A. Darwinismus und Lamarckismus. Munich, 1905.

135. PEARSON, K. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Phil. Trans. (A)*, 1897, **187**.

136. ————. Mathematical contributions to the theory of evolution. On the law of ancestral heredity. *Proc. Roy. Soc. London*, 1898, **62**.

137. ————. On the laws of inheritance in man. I. Inheritance of physical characters. *Biometrika*, 1903, **2**.

138. ————. The law of ancestral heredity. *Biometrika*, 1903, **2**.

139. ————. Mathematical contributions to the theory of evolution. XII. On a generalized theory of alternative inheritance, with special reference to Mendel's laws. *Phil. Trans. (A)*, 1904, **203**.

140. ————. On the inheritance of the mental and moral characters in man, and its comparison with the inheritance of the physical characters. Huxley lecture, 1903. *Trans. Anthrop. Instit. Great Britain & Ireland*, 1903, 179-237.

141. ————. A Mendelian's view of the law of ancestral inheritance. *Biometrika*, 1904, **3**.

142. ————. On the criterion which may serve to test various theories of inheritance. *Proc. Roy. Soc. London*, 1904, **73**.

143. ————. On the laws of inheritance in man. II. On the inheritance of the mental and moral characters in man, and its comparison with the inheritance of the physical characters. *Biometrika*, 1904, **3**.

144. ————. Mathematical contributions to the theory of evolution. On the law of reversion. *Proc. Roy. Soc. London*, 1900, **66**.

145. ————. The grammar of science. (2nd ed.) London, 1900.

146. PEARSON, K., & LEE, A. Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. *Phil. Trans. (A)*, 1901, **195**.

147. PICTET, A. Influence de l'alimentation et de l'humidité sur la variation des papillons. *Mem. Soc. Phys. Hist. nat., Genève*, 1905, **35**; *Nature*, 1905, **72**.

148. POPENOE, P. Twins reared apart. *J. Hered.*, 1922, **13**.

149. POULTON, E. B. Acquired characters. *Nature*, 1894, **51**.

150. ————. Organic selection. *Science*, 1897, **6**.

151. PRZIBRAM, H. Experimental-Zoölogie. III. Phylogenese. Leipzig, 1910.

152. RAYMOND, P. L'hérédité morbide. Paris, 1905.

153. REH, L. Zur frage nach der Vererbung erworbener Eigenschaften. *Biol. Zentbl.*, 1894, **14**.

154. Reibmayr, A. Die Immunisirung der Familien bei erblichen Krankheiten (Tuberculose, Lues, Geistesstörungen). Leipzig and Vienna, 1899.

155. Reid, G. A. Characters, congenital and acquired. *Science,* 1897, **6.**

156. ———. Principles of heredity, with some applications. London, 1905.

157. Rignano, E. Ueber die Vererbung erworbener Eigenschaften. Hypothese einer Zentroepigenese. Leipzig, 1907.

158. ———. Upon the inheritance of acquired characters. Chicago, 1911.

159. Romanes, G. J. Darwin and after Darwin. (3 vols.) London, 1892-1897.

160. Rosanoff, A. J. Dissimilar heredity in mental disease. *Amer. J. Insan.,* 1913, **70.**

161. Rosanoff, A. J., & Orr, F. I. A study of heredity in insanity in the light of the Mendelian theory. *Amer. J. Insan.,* 1911, **48.**

162. Rosenthal, J. Zur Frage der Vererbung erworbener Eigenschaften. *Biol. Zentbl.,* 1889, **9.**

163. Roux, W. Der Kampf der Teile im Organismus. Leipzig, 1881.

164. Ruedin, E. Studien über Vererbung und Entstehung geistiger Störungen. I. Zur Vererbung und Neuentstehung der Dementia Praecox. Berlin, 1916.

165. Ryder, J. A. Proofs of the effects of habitual use in the modification of animal organisms. *Proc. Amer. Phil. Soc.,* 1889, **26.**

166. ———. A physiological hypothesis of heredity and variation. *Amer. Natur.,* 1890, **24.**

167. ———. The inheritance of modifications due to disturbances of the early stages of development, especially in the Japanese domesticated races of gold-carp. *Proc. Acad. Nat. Sci., Philadelphia,* 1893, **45.**

168. ———. Dynamics and evolution. *Biol. Lectures, Woods Hole,* 1894, **2.**

169. ———. A dynamical hypothesis of inheritance. *Biol. Lectures, Woods Hole,* 1895, **3.**

170. Schröder, C. Ueber experimentell erzielte Instinktsvariationen. *Verh. Zool. Ges.,* 1903, **13.**

171. Secérov, S. Die Umwelt des Keimplasmas. II. Der Lichtgenuss im Salamandra-Körper. *Arch. f. Entwickmech.,* 1912, **33.**

172. Semon, R. Die Mneme als erhaltendes Princip im Wechsel des organischen Geschehens. *Biol. Zentbl.,* 1905, **25.**

173. ———. Ueber der Erblichkeit der Tagesperiode. *Biol. Zentbl.,* 1905, **25.**

174. ———. Beweise für die Vererbung erworbener Eigenschaften; ein Beitrag zur Kritik der Keimplasmatheorie. *Arch. f. Rassen- u. Gesellschbiol.,* 1907, **4.**

175 ———. Die Abstammungslehre. XV. Können erworbene Eigenschaften vererbt werden? Jena, 1911.

176. ———. Das Problem der Vererbung erworbener Eigenschaften. Leipzig, 1912.

177. Siemens, H. W. Die Zwillingspathologie. Berlin, 1924.

178. ———. The diagnosis of identity in twins. *J. Hered.,* 1927, **18.**

179. Sioli, —. Ueber directe Vererbung von Geisteskrankheiten. *Arch. f. Psychol.,* 1885, **16.**

180. Snyder, L. H. The inheritance of the blood groups. *Genetics,* 1924, **9.**

181. ———. Human blood groups and their bearing on racial relationships. *Proc. Nat. Acad. Sci.,* 1925, **11.**

182. ———. Human blood groups: their inheritance and racial significance. *Amer. J. Phys. Anthrop.,* 1926, **9.**

183. Sollier, —. Du rôle de l'hérédité dans l'alcoholisme. Paris, 1889.

184. Sommer, M. Die Brown-Séquardische Meerschweinschenepilepsie und ihre erbliche Uebertragung auf die Nachkommen. *Beitr. path. Anat. allg. Path.*, 1900, **27**. P. 46. Summary by Ziegler, *Zool. Zentbl.*, 1900, **7**.

185. Sommer, R. Familienforschung und Vererbungslehre. Leipzig, 1907.

186. Spencer, H. Principles of biology. (1st ed., 2 vols.) 1864-66. (2nd ed. of Vol. I, 1899.)

187. ————. The inheritance of acquired characters. *Nature*, 1890, **41**.

188. ————. The inadequacy of natural selection. *Contemp. Rev.*, 1893, **63**

189. ————. Professor Weismann's theories. *Contemp. Rev.*, 1893, **63**.

190. ————. A rejoinder to Professor Weismann. *Contemp. Rev.*, 1893, **63**.

191. ————. Weismannism once more. *Contemp. Rev.*, 1894, **64**.

192. Sprott, W. J. H. Physique and character. (Tr. from E. Kretschmer.) New York, 1925.

193. Stockard, C. R. Experimental modification of the germ-plasm and its bearing on the inheritance of acquired characters. *Proc. Amer. Phil. Soc.*, 1923, **62**.

194. Stockard, C. R., & Papanicolaou, G. A further analysis of the hereditary transmission of degeneracy and deformities by the descendants of alcoholized mammals. I, II. *Amer. Natur.*, 1916, **50**.

195. ————. Further studies on the modification of the germ-cells in mammals: The effect of alcohol on treated guinea pigs and their descendants. *J. Exper. Zoöl.*, 1918, **26**.

196. Szily, A. von. Die Ontogenese der idiotypischen Spaltbildungen des Auges, des Mikrophthalmus und der Orbitalcysten. *Zsch. f. Anat. u. Entwickgesch.*, 1924, **74**.

197. Thomson, J. A. Heredity. (2nd ed.) London, 1908. New York, 1910.

198. Vicari, E. M. The non-inheritance of the effects of training. *Science*, 1924, **59**.

198*a*. Ward, J. Heredity and memory. Cambridge, 1913.

198*b*. Watson, J. B. Behavior: an introduction to comparative psychology. New York, 1914.

199. Weismann, A. Die Kontinuität des Keimplasmas, als Grundlage einer Theorie der Vererbung. Jena, 1885.

200. ————. Zur Frage nach der Vererbung erworbener Eigenschaften. *Biol. Zentbl.*, 1886, **6**.

201. ————. Ueber die Hypothese einer Vererbung von Verletzungen. *Naturf. Versamml. in Köln*, 1888.

202. ————. Essays upon heredity and kindred subjects. Oxford, 1891, 1892.

203. ————. The germ-plasm: a theory of heredity. London, 1893.

204. ————. The effect of external influences on development. Romanes lecture. Oxford, 1894.

205. ————. On germinal selection as a source of definite variation. Chicago, 1896. (2nd ed., 1902.)

206. ————. The evolution theory. London, 1904.

207. ————. Richard Semon's "Mneme" und die "Vererbung erworbener Eigenschaften." *Arch. f. Rassen- u. Gesellschbiol.*, 1906, **3**.

208. Wettstein, R. von. Der Neo-Larmarckismus und seine Beziehungen zum Darwinismus. Jena, 1903.

209. Wheeler, W. M. The physiognomy of insects. *Quar. Rev. Biol.*, 1927, **2**.

210. Wiggam, A. E. Twins again. *J. Hered.*, 1923, **14**.

211. Wilder, H. H. Physical correspondence in two sets of duplicate twins. *J. Hered.,* 1919, **10**.

212. Windle, B. C. A. Teratological evidence as to the heredity of acquired conditions. *J. Linn. Soc.,* **23**. *Nature,* 1891, **40**.

213. ————. A note on identical malformations in twins. *J. Anat. & Physiol.,* 1892, **26**.

214. Woods, F. A. Twins prove the importance of chromosomes. *J. Hered.,* 1919, **10**.

215. Zeleny, C. The relative numbers of twins and triplets. *Science,* 1921, **53**.

216. Ziegler, E. Können erworbene pathologische Eigenschaften vererbt werden, und die entstehen erbliche Krankheiten und Missbildungen? Jena, 1886.

217. Ziegler, H. E. Die Vererbungslehre in der Biologie. Jena, 1905.

CHAPTER 2

THE STUDY OF LIVING ORGANISMS

W. J. Crozier
Harvard University

> "Scientific psychology is a part of physics, or the study of nature; it is the record of how animals act. Literary psychology is the art of imagining how they feel and think."
>
> Santayana, *Scepticism and Animal Faith*

I

It is generally agreed that the interpretation of the properties of living organisms is exposed to hazards of more varied and insidious sorts than those attending the scientific treatment of non-living objects. By science one understands the unresting attempt to obtain a comprehensive, satisfying account of the universe and all that it contains; the essence of this account, as science, is that it must be based upon the fewest possible assumptions. Why such an attempt should be found satisfying, presents a problem into which we need not now inquire. It has been most successfully pursued in the realm of physics. Until about 1800 physics had been concerned with the formulation of a conception of phenomena as they might be described in terms of "forces" acting between one body and another. More recently it has come to deal with ideas of a higher order of complexity, according to which the energy of a material system is regarded as controlled by the configuration and motion of the system, and it is thus concerned with aspects of physical systems transcending any complete analysis of details.

The biological system presented by a single individual, no less than that which may be conceived to be presented by any association or society of different individuals (136-138), is obviously not a "thing," a single event, but a system of *relations*. It is these relations which must be defined through investigation. It would appear to be useless to quarrel with the idea that to do so requires a procedure essentially mathematical, because quantitative; although not so much concerned with that department of mathematics dealing with the enumeration of numbers but, on the contrary, based upon the notion of functional dependence. This view is not alone attractive because of its historical success in the case of physics, but is easily supported by appeal to ordinary sense and the judgment of authority. Its greatest force is obtained from the common-sense logic of the definition of units. "The most important step in the progress of every science is the measurement of quantities. Those whose curiosity is satisfied with observing what happens have occasionally done service by directing the attention of others to the phenomena they have seen but it is to those who endeavor to find out how much there is of anything that we owe all the great advances in our knowledge" (109*a*, Chap. IV). The measurement of quantities, which is necessary for the production of statements of functional dependence, implies units of measurement possessing definite dimensions, using "dimensions" in the physical sense.

From the most general standpoint, everything which a living organism does, including, for example, its development, might be termed an aspect of its "behavior"; the trouble with such an attitude is that it leads to vagueness, and the several attempts which have been made at the discussion of the conduct of organisms on this basis have not been exactly fruitful. Nor can it be held that the employment of incompletely worked-out analogies between properties of non-living systems and of organic conduct, such as one frequently discovers in discussions of the "emergent" quality of organisms and particularly of the configurationist modes of interpretation, is either satisfying or in any sense intrinsically very productive. What is required is formulation of conduct in such a way as to permit measurable prediction of behavior. It is unfortunately true that many persons, no matter how intelligent, logically acute, and imaginatively alive they may be, are frequently paralyzed by the sight of a mathematical symbol. It has been the case in biology that the complexity of its phenomena has on the whole tended to encourage the emotional bias of this type of worker. It is sufficient to point out, however, that it would appear to be an inescapable phase in the development of biology that there must ensue a long period of at first more or less unsuccessful attempts to obtain functional formulations of the properties, particularly of the behavior, of organisms before it will be worth while to discuss the question as to whether non-physical attributes must be assigned to them.

It is sometimes supposed that the changing viewpoint in modern physics, altering the conception of nineteenth-century mechanics, is unfavorable to those who have worked for a "mechanistic" interpretation of vital processes; the efficient reply is that the aim of mechanistic interpretation is to attain a basis for the common treatment of living and of non-living objects. The emotional rebellion at the thought that man may be a machine is largely based upon an inadequate conception of what a machine may be, nor does the development of now outworn notions about physical machines imply in any proper sense that these must be granted organic or quasi-biological properties. For this reason, one need not devote much time to the rather stupid quarrel supported by those who would stress the "non-mechanical," by implication non-mathematical, properties of organisms. The behavior of organisms presents from this standpoint what is unquestionably the most difficult of biological themes. This is in one sense its major attraction. If we consider for a moment the parallel provided by the state of theory dealing with facts of inheritance as it existed previous to 1900, and compare this with the solid, rational, theoretic structure of genetics which has been reared in the short time elapsing since then, on the basis of what is essentially a simple mathematical conception (116) embodied in a theory of the gene which is in effect independent of any chromosomal or other crude mechanism, it is easily seen that there is abundant ground for encouragement.

Such advances are at bottom the result of a persisting tendency to extend to the most complex phenomena the general principle of the uniformity of nature and of the continuity of physical laws in space and time. The deduction of the composition and physical state in the interior of a distant star would appear to present no more serious a logical difficulty than does the interpretation of the behavior of an organism. In fact, there are certain instructive similarities. In the case of a distant star the analysis must be made by indirect methods, since the material of which it is composed is not directly open to experimental handling. With an organism the properties which most concern us are more than likely to be obliterated as the result of any attempt at gross interference—again, the requisite analysis must proceed by somewhat indirect channels. The checking of astrophysical deductions, the results of "boring into distant stars," can be done by testing the consequences of the mathematical implications; as yet, no monumental failure has attended this procedure, which involves (again as in the case of an organism) descriptions of systems which are altering as a function of time. In a way which is entirely similar, the results of indirect analysis of the conduct of organisms are also to be checked by examination of the mathematical implications of the treatment.

Historically and, as it would appear, largely owing to the association of academic philosophy with psychology, the development of such an attitude has been hampered by the impulse to work from a "general," all-inclusive, theoretical interpretation, before data making an interpretation possible have been obtained. To secure such data is a primary task of the experimental investigation of conduct. At one time it was supposed that this undertaking would proceed most easily on the basis of experiment with the allegedly simpler unicellular (better, non-cellular) forms, in which "elements" of conduct in general might be recognizable. But it is easily seen that the investigation of irritability and transmission ought to be most productive when one is able to deal with living structures in which these particular activities are available in the most unobscured, isolated, best-developed ways. It is upon the results of such investigations that it becomes possible to construct an increasingly more and more complete picture of the events which succeed one another in the chain of disturbances initiated in a sense-organ, transmitted over nervous or nerve-like pathways, entering into competition with other similar disturbances in the central nervous system of higher animals, and ultimately leading to the release of energy in muscles, glands of various kinds, and other effector parts. The attempt to obtain laws of behavior, however, which shall be descriptive of behavior as such can probably not be obtained with sufficient completeness by this particular procedure. It is specifically with this problem that we shall be concerned.

II

A monograph by Loeb (104) summarized his position with regard to the chief evidence then available concerning tropistic conduct and the deductions which might follow from its analysis. Loeb's approach to the problem of individual organic behavior was initially and primarily philosophic, in the sense that through recognition of the elements of action and the modes of their control he looked for an answer to the puzzles of free will and the determinism of conduct. Such elements he found in the machine-like orientations of animals. With the insight permitted to a superbly penetrating intelligence, he was able to see beyond the immediate consequences of relatively simple experiments and to extract from them their utmost general significance.

The primary assumption of the tropism doctrine is that, in instances properly regarded as falling under this description, the enforced movements resulting in orientation cease when the organism is equally affected upon its two symmetrical sides (100, 101, 104). Such movements are regarded as *forced,* in the most general sense, because they are inevitably results of the structure of the reacting organism and of the presence of appropriate substances in its sense-organs (or in one or more areas serving as loci for excitation) symmetrically arranged. The term "forced movement" was at first introduced into such discussions from brain physiology (104). The wider understanding of the underlying notion, however, more easily permits embracing relevant features of the behavior of plants and lower organisms as well as those in certain higher animals within the scope of a single interpretation. Dynamically, the circling motions of a dogfish forced to swim in spirals to the right when its left lower optic lobe has been removed are entirely comparable to the predominantly right-hand direction of swimming (124) in an uninjured dogfish with its left nasal sac plugged with cotton and in the presence of juices diffusing from crushed meat. The tonus of swimming muscles on one side has in the first case been lowered by interruption of nervous pathways, and on the heteronomous side reflexly increased; in the second case the difference in tonus has been brought about by continuous unilateral excitation.

A large number of instances of this kind are now recognized,[1] and through repetition in texts some of these have become rather shopworn. We may therefore cite several of more recent date. A quietly creeping meal-worm (larva of *Tenebrio*), or a diplopod such as *Spirobolus,* responds to the impress of unsymmetrical tensions on its musculature, as when the body is bent in an arc, by a prompt movement of the anterior end which, as with an earthworm similarly treated (117), results in the alignment of the axis of this end parallel to that of the posterior part (114, 67, 48, 49). Contact with a lateral surface inhibits this response (Figure 1). The balanced action of equal zones of contact on either side results in the pursuit of a straight course (Figure 2); similar effects are obtainable in other forms, for example, in young mice and rats (Figure 3), and may be taken as illustrating the manner in which asymmetric excitation results in forced redistributions of tension of orienting structures, these ceasing when excitation has become equalized on the two sides.

---

[1] Cf. (22, 12, 8, 122, 129a) for convenient bibliographic citations, and the detailed annual reviews in (14).

The theory of the phototropic movements of insects has regarded the forced orientations under illumination as due to the effect of asymmetric sensory excitation upon the postures of opposed appendages (93, 10, 83). It is not always clear to what extent the frequencies of phasic limb movements (39), as well as the postural influence upon effective lever action

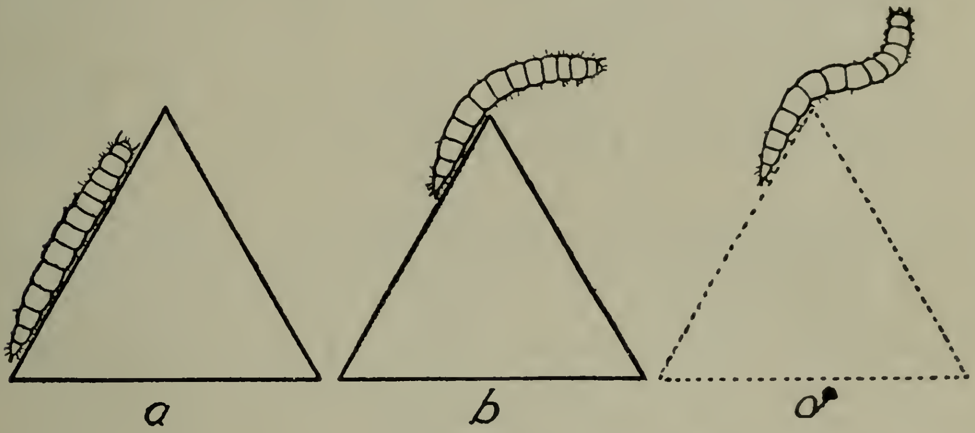


FIGURE 1

Stereotropic Orientation of Larval *Tenebrio* (*a, b,* Successive Positions) and Reflex Homostrophic Orientation Released (at *c*) when Source of Contacts is Removed

(From *J. Gen. Physiol.*)

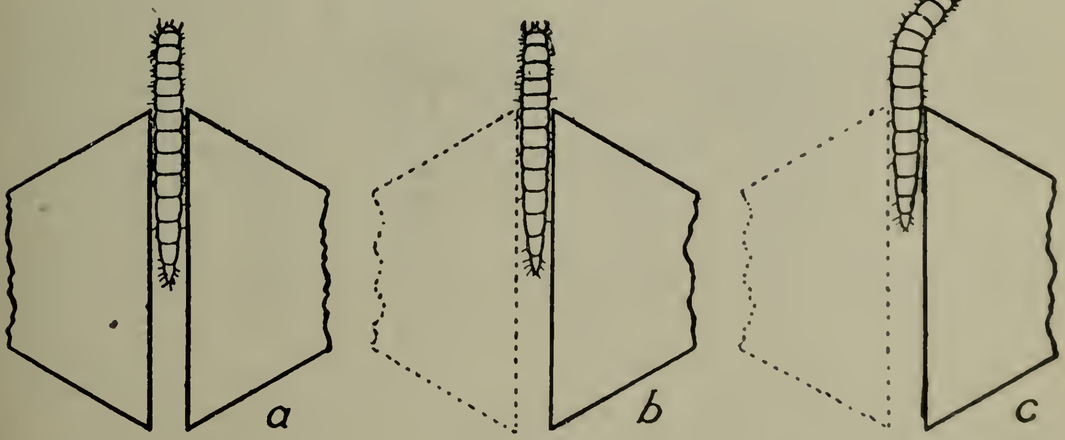


FIGURE 2*a*

Balanced Action of Equal Bilateral Contact Shown in Pursuit of a Straight Course (*a*)

Removal of contact plate on one side (*b*) is immediately followed by stereotropic bending (*c*). (From *J. Gen. Physiol.*)

of the legs in progression, have been separated as concerned in the turning movements (107, 108, 30, 31). However, in sufficiently uncomplicated cases the rôle of posture is transparent. If one eye of the water scorpion *Ranatra* be removed (or covered), the insect under water continues for a long period to circle toward the remaining eye if it is illuminated from above. During creeping, in air or on vegetation under water, the opposite legs of each locomotor pair are used alternately, the diagonally opposite legs being in phase; but in swimming under water the legs of a morphological pair work synchronously, in phase, the two pairs alternating in the execution of the propulsive stroke. If both eyes have been covered or removed, and the animal put in a large cylindrical tank illuminated from a single source of light directly above, the *Ranatra* stays near the surface and moves irregularly. But if only one eye has been prevented from receiving excitation, the positively phototropic swimming, due to the greater extension of the legs upon the blind side, results in circus movements toward the side of the functional eye (Figures 4, 5).

It should not be supposed that emphasis upon activities of this type seeks to imply that all the movements of organisms are tropistically regulated. A point which will bear restating is the distinction Loeb drew between *phototropic* movements and responses to change of intensity of light (22); the latter, leading, for example, in the case of motile forms reacting negatively to increase of illumination, to aggregation in relatively shaded areas, has nothing to do with the phototropism, positive or negative, which they may also exhibit at the same time (57, 29). In many cases major features of the behavior of animals, involving actions of prime significance for their perpetuation, are obviously consistent with modes of tropistic response which experiment under controlled conditions may reveal, but

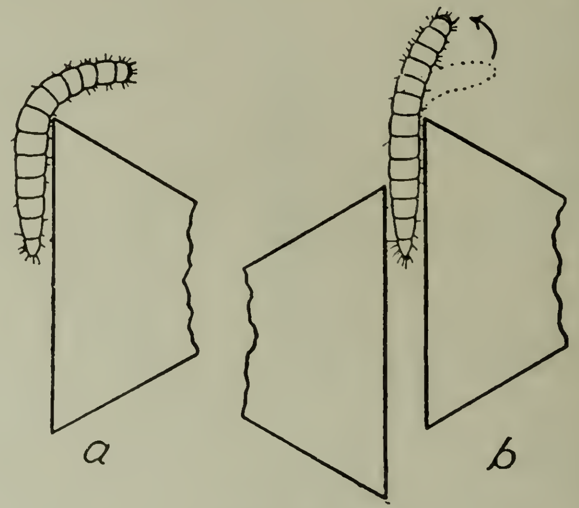


FIGURE 2*b*

A larva in stereotropic orientation toward unilateral contact (*a*) responds by lesser stereotropic curvature when a less extensive contact is introduced at the other side (*b*). (From *J. Gen. Physiol.*)

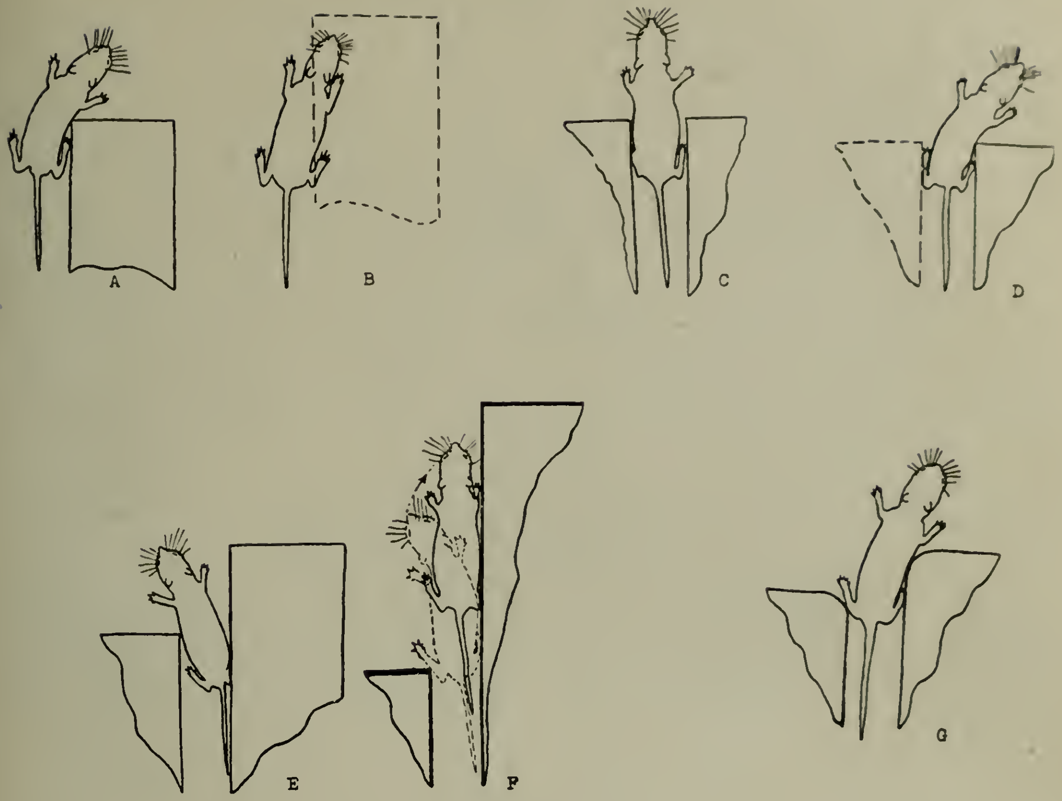


FIGURE 3

(A) Stereotropic orientation of young rat or mouse at the corner of a box along one side of which it has been creeping.

(B) A young rat or mouse has been creeping in contact with the side of a box (dashed outline); the removal of the box results in partial orientation toward that side.

(C) A young rat or mouse creeping in a passage-way between two boxes, just wide enough to permit gentle contact on either side during the animal's swaying progression, is found to emerge from the passage-way without orientation. Equivalent bilateral stimulations prevent stereotropic turning.

(D) An individual emerging from equal bilateral contacts with two boxes proceeds in a straight path, without orientation; but if one of the boxes be removed (dashed ouline), it promptly orients toward the remaining one.

(E) Contact at one side with the corner of a box may lead to orientation toward that side, apparently due, in part at least, to more intense tactile excitation than is provided by a continuous flat surface (or by smoothly rounded corners).

(F) When such a corner is passed, orientation persists toward a continuing contact on the opposite side.

(G) When blocks providing lateral contacts are of unequal extent, the young rat or mouse orients toward the side of more extensive contacts, but does not completely turn the corner unless the difference in extent of the two blocks is more than half the length of the animal. This, the expected result from a tropistic standpoint, is obtained when the corners of the contact blocks are smoothly rounded. (From *J. Gen. Physiol.*)

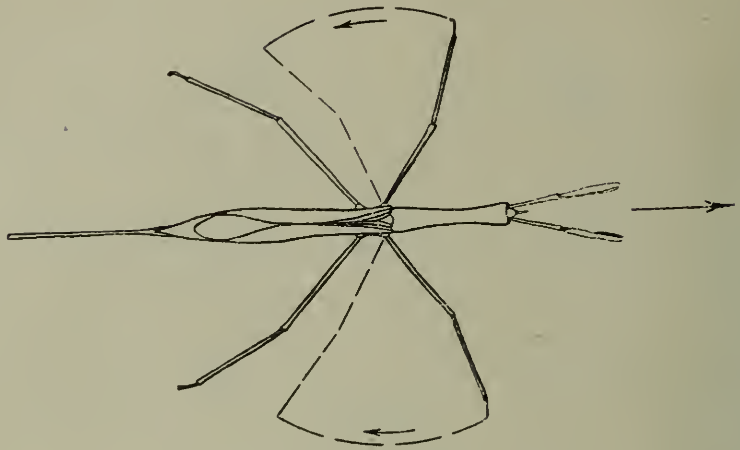


FIGURE 4

Showing Positions of the Locomotor Appendages in a Swimming *Ranatra* from which Both Eyes Have Been Removed

The two limbs of a pair move synchronously and in phase. At the conclusion of the stroke (dashed lines) the legs are slightly more extended than at the beginning of the stroke (full lines); as the beat of the anterior legs is concluded, the posterior pair is moved to a position comparable to the initial position of the anterior limbs; thus the two pairs move in alternate phase. (From *J. Gen. Physiol.*)

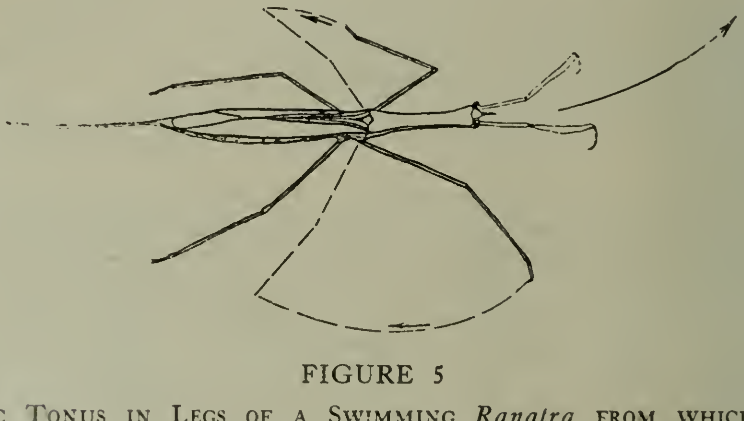


FIGURE 5

Asymmetric Tonus in Legs of a Swimming *Ranatra* from which One Eye (Right) Has Been Removed

The legs on the blind side assume positions comparable to those of the appendages of a completely de-eyed *Ranatra* (Figure 4). On the eye side the tonus of leg flexors is increased. This results in hydrostatic tilting of the whole body at rest. In swimming, the two legs of a pair move synchronously but the greater flexor tonus on the eye side is maintained throughout the execution of the swimming stroke, so that the extension of the left legs during the propulsive motion (arrows) is less than on the relatively atonic right side. The animal accordingly swims continuously in a circle toward the eye-bearing side. (From *J. Gen. Physiol.*)

it is also true that curious instances are encountered in which there exists superficially a striking exception to this state of affairs. The behavior of *Onchidium floridanum,* for example, at Bermuda exhibits a number of arresting features (58, 29). This is a small shell-less pulmonate which lives in holes and crevices of the limestone shore between tidal limits. Each inhabited cavity shelters, when the tide is in, a group of the mollusks. As low water approaches, and if the weather is fair, but only once in the twenty-four hours and only during daylight, the animals emerge one at a time from the "nest" and wander for limited distances over the algal carpet of the rocks, on which they feed. Before the tide rises again, the individuals emanating from any one nest simultaneously return to that nest. Certain possible factors in the control of this "homing" behavior have been discussed. In the laboratory or on an artificial surface in the field *Onchidium floridanum* is negatively heliotropic. It also reacts definitely to quick reduction of the intensity of light. The tentacular eyes are of no significance for these responses, which are mediated by integumentary receptors. Yet the mollusks emerge from their crannies only in the daytime, and in their natural movements on the shore they are in no sense guided or directed by the light. It was possible to show that the phototropic responses are completely inhibited during the creeping of these animals upon the rocks immediately surrounding the specific home site. Removed to a new location, whether inhabited or not by other colonies of the same form, *O. floridanum* is at the mercy of its heliotropic response and of its retractive response when shaded; it is thereby prevented from finding a new "home" before the tide rises, and becomes washed away. In Cuba, upon the other hand, another species of *Onchidium* has been found (68) to occur only in shaded places, and to be free from the necessity of returning to its specific home, although otherwise its behavior is similar; the natural movements of this form on the shore are entirely consistent with its negative heliotropism.

Nor does the fact that tropistic reactions may be modified through experience at all alter their significance, remove their mechanical nature, or destroy their utility for analytical purposes when they are exhibited. The modification or, as in the case of *Onchidium,* the suppression of such responses presents merely an inviting opportunity to discover the molecular mechanisms whereby the modification may be produced. The chemical control of phototropic movements and of their positive and negative character (22) remains to be developed as a means to this end. A quantitative treatment of tropistic behavior is essential if the understanding of conduct is to be furthered in this way. By "quantitative" one does not wish to imply simply that which has to do either with numbers of observations or of animals, nor yet merely numerical data which may be submitted to statistical treatment. One does imply the discussion of numerical data which may be accorded analysis in terms of intelligible theory; the "quantitative" has significance in the light of functional interrelationships which the analysis may reveal and of the tests which it suggests.

From this standpoint the recognition and the characterization of a particular tropism in a specified organism is scarcely an exciting end in its own right. The tropism in question may become, however, a powerful instrument of analysis under suitable conditions. If it can be shown, for example, that in controlled circumstances the direction of movement of given animals may be predictably determined even when more than one orienting force is at work, we then have at hand a means (as yet in an elementary stage of development) whereby a second-order test of theories of oriented movement may be applied. Moreover, this general procedure may possibly be employed to obtain information about central nervous states and conditions not otherwise approachable; the mechanism whereby the central nervous system of a given animal functions in the adjustment of the effects of competing modes of excitation, originating in different peripheral areas, must be undertaken by some such method.

III

If the tropistic behavior of animals is in any real degree that of orienting machines, it should be possible to deal with such problems as the following: knowing the orienting efficiency of light of a particular composition, from a single source, it should be possible to predict the form and amount of the orientation when lights from two similar sources play upon the animal. The question is but deceptively simple, yet in certain essentials it can be answered. The answer is based chiefly upon the single assumption of the tropism concept and involves quantitative knowledge of the connection between light intensity and orienting efficiency, with certain implicit assumptions (which can be tested) regarding the central nervous interplay between excitations of different peripheral origins.

One detects the prevalence of two general bodies of opinion regarding tropisms. If we may somewhat harshly overstate the case, one type of opinion seems to regard the whole matter as settled; and the other, as being of no moment. It would be difficult to determine which view falls wider of the mark. Even in the comprehension of the problem, so far as this may be revealed in definitions explicit or implied, one constantly encounters curiosities. A tropism is most effectively defined as an oriented movement in an energy field.

When the exciting field is a luminous field, part of the desired information may be had from circus movements. In the slug *Limax maximus* the eyes are the effective photoreceptors. With illumination from above, a dark-adapted *Limax* from which one eye has been removed circles continuously toward the non-stimulated side. The circling posture is continuously maintained until photic adaptation supervenes. The posture of the body is due to differences in the maintained tensions of the muscles of the two sides of the body-wall. The tension difference is maintained only under illumination, and disappears in darkness. With proper care to insure absence of air currents, odorous stimulation, and the like, and with proper selection of phototropic individuals, a measure of the orienting posture is easily secured. The experiment is particularly significant because the muscular mechanisms for creeping and for turning are structurally distinct in such gastropods (29, 68). It is necessary, however, to take account of the fact that the two are nervously interrelated in such a way that faster progression (controlled by the release of pedal waves) involves less effective turning. The two mechanisms can be separated in an analytical way through their relations to temperature (62, 63) and in other ways. A large number of such experiments has been made by Cole and Crozier, in which the circling trails could be accurately recorded and measured. With comparable speeds of progression, the phototropic effect is measured by the ratio *degrees turned per centimeter of path.*

While it might seem to be a matter of simple common sense, it is nevertheless necessary to emphasize that for such experiments selection of individual animals is just as necessary as control of light intensity and of temperature. A confusion unquestionably exists in some quarters as to the objective of such experiments; it seems to be supposed that the purpose is to examine a random lot of individuals of a particular sort and to note whether they are phototropic. This kind of inquiry can result, at best, in statements to the effect that under such and such conditions *x per cent* of the organisms were seen to respond; which tells us precisely nothing about the character or the measurable attributes of the response when it is exhibited. At the moment we are concerned to discover the relation between light intensity and orienting effect in phototropic *Limax,* and non-reactive individuals [such as are easily produced at will by certain experimental treatments, (65)] are of no interest.

It turns out that the circus movement effect, expressed as the amount of turning per unit length of path, is directly proportional to the logarithm of the light intensity, over a fairly wide range (Figures 6, 7). This kind of relation is often labelled "Weber's law," apparently with the thought that it is thereby explained. The status of the Weber-Fechner formulation in relation to photic excitation has been adequately treated by Hecht (85-90); certain more general considerations will be referred to

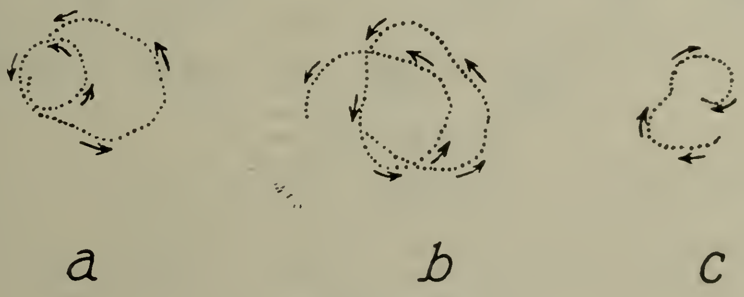


FIGURE 6

Typical Circus Movement Trails

a, at 21°, 18.73° per cm. deflection, 7.94 cm. per min., speed of creeping; *b,* at 21°, 15.53° per cm., 9.42 cm. per min.; *c* at 27°, 30.45° per cm., 5.08 cm. per min. Note continuous character of the circling motion. There is no difference, quantitatively, between right-hand and left-hand turns. Scale 1:4. (From *J. Gen. Physiol.*)

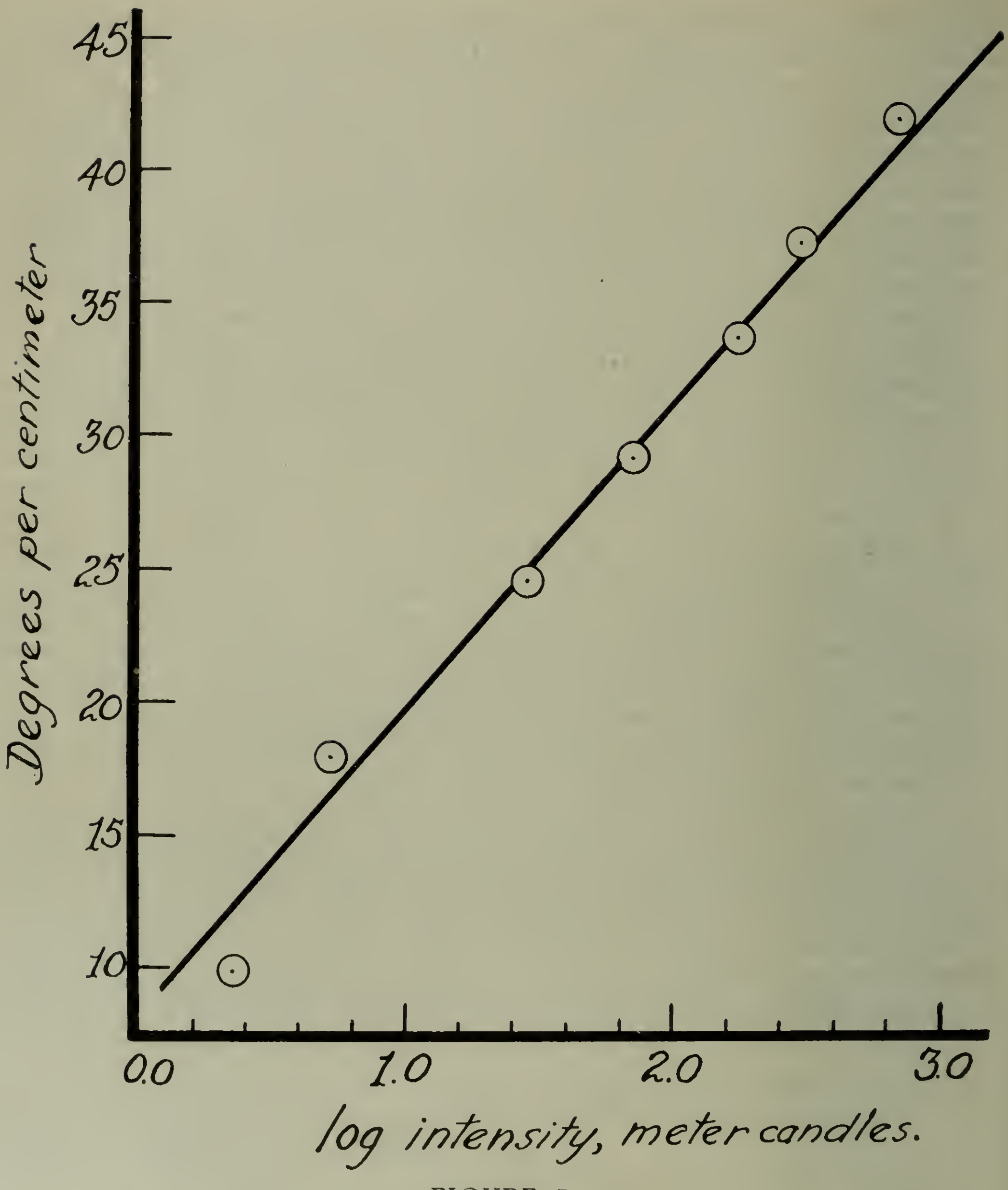


FIGURE 7

Data from Experiments by Crozier and Cole on the Circus Movements of the Slug *Limax* (Negatively Phototropic)

One eye-bearing tentacle has been amputated; illumination is from directly above; the measure of circling posture is given by *angular deflection per unit length of path;* mean values of this quantity are plotted against corresponding values of log *I*, where *I* is the intensity of illumination at the photoreceptive surface of the remaining tentacle. Light-adaptation being slow, the circling movement should continue for some time, which is the fact; and the tropistic effect should be independent of the animal's position, inasmuch as only one eye is functioning. The effect is expressed as a difference in tension reflexly established on the two sides of the body, continuously maintained during continuous excitation. This postural effect is proportional in magnitude, over the range of illuminations used, to log *I*.

subsequently. What is important is to realize that such empirical relationships may be extremely useful, where formulations of greater theoretic implication are clumsy or unmanageable. There are other methods whereby the orienting power of light from a single source may be expressed (141, 79, 142), but their explanation involves additional considerations. In general, however, they lead to two forms of relation in different cases: the orienting power is directly proportional to the light intensity or, more commonly, to its logarithm. Both relations are consistent with elementary photochemistry, and with the fact that for the production of a given (measured) photosensory effect with light of a given intensity the Roscoe-Bunsen law (or Talbot's law) is adequately obeyed (104, 86, 87, 121).

When circus movement effects are studied under proper conditions (39, 62, 63), the illumination of the organism does not change during orientation, and the resulting data should permit one to predict the outcome when a phototropic organism is illuminated from one side. In a number of cases in which it has been attempted to investigate circus movements on a horizontal surface with horizontal illumination, it has been recorded that a phototropic animal with one eye covered at first performs

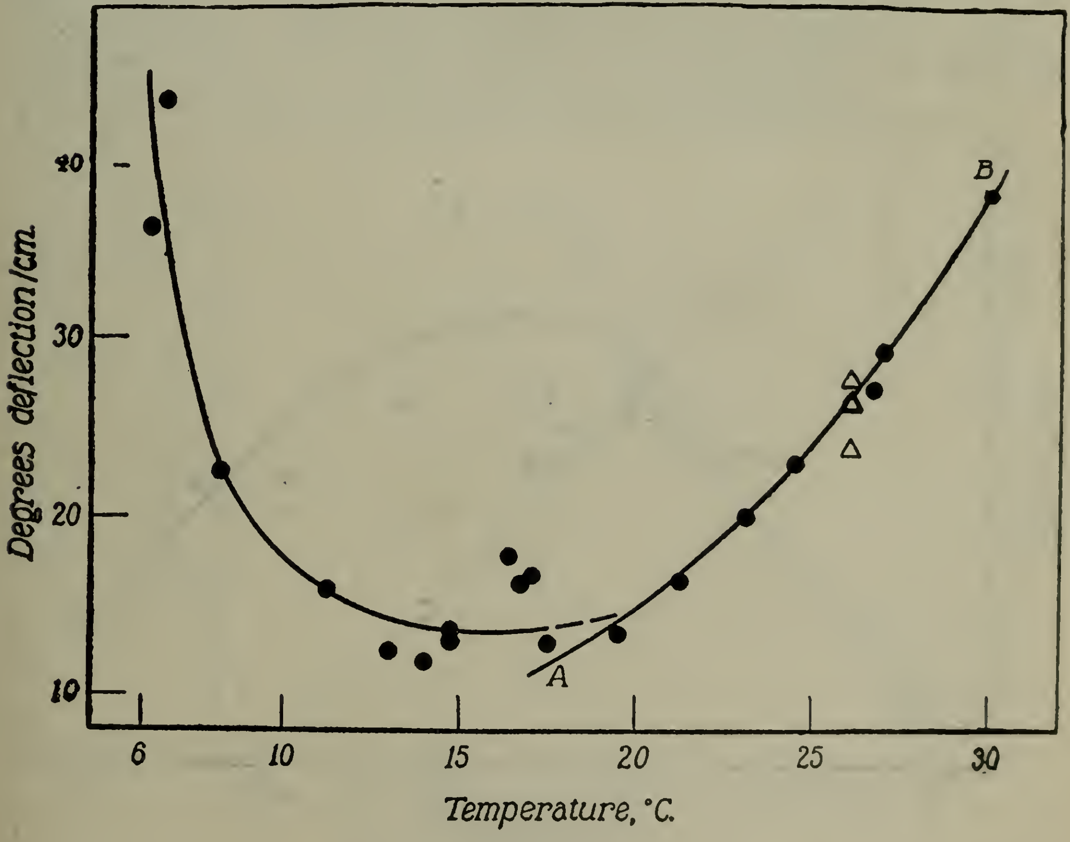


FIGURE 8

The Ratio *Degrees Deflection: Cm. Length of Path,* in Circus Movements under Constant Vertical Illimination, as a Function of Temperature

Several independent determinations shown as triangles. The curve *A-B* is a theoretical one, such that μ = 16,800. (From *J. Gen. Physiol.*)

circus movements in creeping toward the light, but that after successive trials it may come to creep directly toward the light. It has been suggested that this phenomenon simulates learning, and might be due to the gradual elimination of motions tending to carry the animal away from the light, and thus to a kind of psychic effect connected with repeated excursions toward the light. It has been shown, however, that under these conditions there intervenes a simple photic adaptation bringing about a lessened effectiveness of the excitation upon the functional eye (71, 38). But it is difficult at present to use these facts in a quantitative way. The case is sometimes complicated by the kind of effect which has led to von Buddenbrock's (35, 36) conception of *Lichtkompassbewegung,* namely, that since a positively phototropic caterpillar may approach a small source of light in its own plane along a *spiral* path even with one or more turnings about the light, it is really responding in such a way as to put its body axis at a constant angle to the light rays reaching it from moment to moment. This behavior has been carefully examined by Crozier and Stier in larvae of *Malacosoma* (tent-caterpillars). A path such as von Buddenbrock requires would be a logarithmic spiral, which has well-known simple properties. The actual paths do not exhibit these properties. Therefore the *Lichtkompass* effect need not be invoked in such cases, and indeed is no explanation. The spiral paths are in fact due to the simultaneous ex-

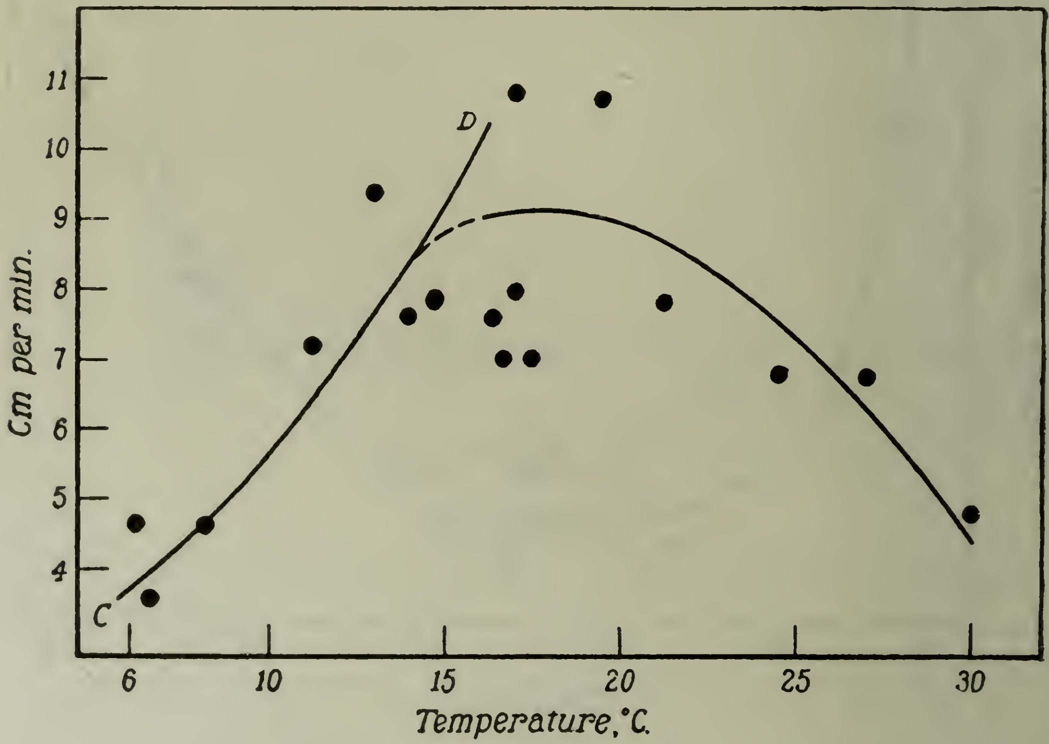


FIGURE 9

The Rate of Creeping during Circus Movements under Constant Illumination at Different Temperatures

(From *J. Gen. Physiol.*)

hibition of two separately identifiable modes of response: the orienting effect of the light increases the tonus of longitudinal muscles on the illuminated side, and the bending of the caterpillar's body so produced evokes the body-straightening movements which constitute the homostrophic reflex (117, 114, 115, 67, 48, 49). The animal consequently proceeds under the influence of two directive vectors—one, increasingly preponderant with closer approach to the light, directed to the light and varying as the inverse square of the distance (or its logarithm); the other, leading by itself to tangential motion with constant speed, is expressible as of constant magnitude. The actual path represents a continuous adjustment between these two forces, and trails of the kind found are easily obtained by graphical construction on the assumption that they are the efficient controlling elements. It is of interest, none the less, that the equations of such loci, for all the simplicity of the assumptions, are very difficult to obtain. But this is purely a matter of mathematical complexity, and cannot fairly be charged to the unpredictability of the conduct of the caterpillars. There is illustrated, however, one kind of reason, often overlooked, why analytical experimentation must involve rigidly simplified conditions; it is true in this field as in others, as Boscovitch long ago suspected, that it is often not so much a question of the complexity of nature as of the deficiency of our mathematics.

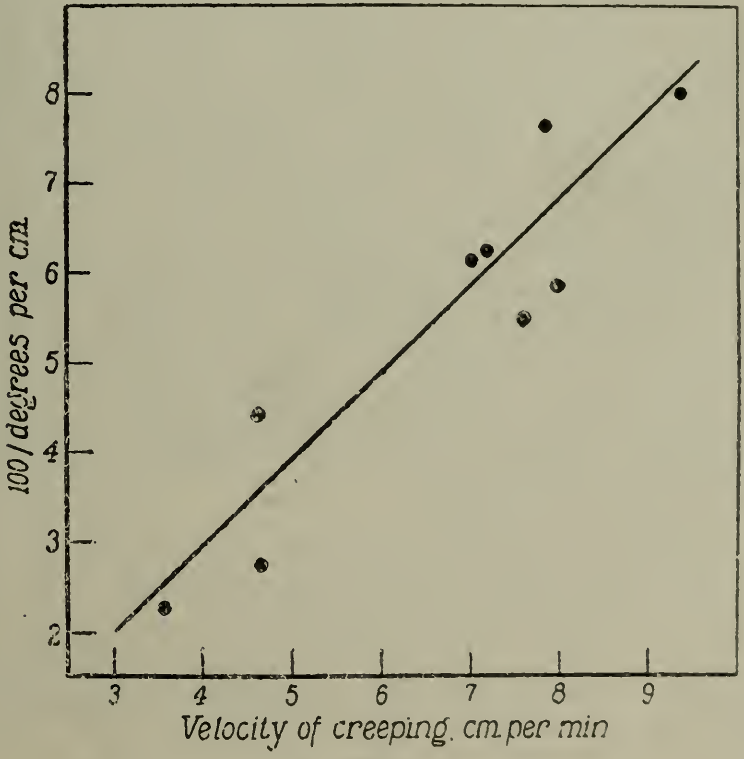


FIGURE 10

At Temperatures up to About 15° the Reciprocal of the Amplitude of Turning Movements is Directly Proportional to the Velocity of Creeping

(From *J. Gen. Physiol.*)

The point here raised is intimately connected with another. It is sometimes voiced as a reproach that simplification requires "unnatural conditions"—whatever "unnatural" may be supposed to mean. An instance is supplied by the phototropic circus movements of *Limax*. Experiments over a range of temperature show that in the performance of circus movements under light of constant intensity two chief factors are involved—the rate of creeping and the amplitude of turning. The structural basis for each is distinct, but the two are functionally interrelated in various ways. Within a zone of temperatures in the immediate neighborhood of 16°-15° C., which probably would be commonly accepted as "normal" for this animal, the response is distinctly more variable, quantitatively, than at temperatures either higher or lower. Above 16° the amount of turning increases with rise of temperature, while below 15° the reverse is true (Figures 8, 9, 10). In this region of "normal" temperature conditions, the balance between (1) the tendency to release pedal waves, which result in progression, and (2) the turning tendency, is close to a critical point (for related phenomena involving discontinuities in the effects of temperature, cf. 50, 52). Slight circumstances may thus determine in particular instances which of the two effects is dominant—hence statistical variation in the amplitude of turning per unit length of path. At higher temperatures the turning mechanism is in control, and the slug creeps only so fast as the central mechanism having to do with differential bilateral tonus will permit. At lower temperatures the amount of turning is limited and controlled by the speed of creeping, to which it is inversely proportional. The "control" of creeping by the central nervous determination of differential tonus in the musculature of the body is demonstrated by the results of treatment with strychnine (63), which suppresses phototropism.

Under "natural" conditions, consequently, certain kinds of variability in response may be fairly conceived to result from the fact of delicate dynamic adjustment among processes which (in effect) compete for control of the animal's mechanisms of movement. The perhaps curious principle ensues that, for real understanding of the determination of conduct and the identification of its elements, it is practically essential to place the organism under distinctly "abnormal" conditions such as are favorable to the unhampered expression of separate elements of the usual nexus. From a slightly different standpoint the behavior of *Onchidium* is illuminating in this respect (58, 29).

In the examination of phototropism it is found possible to account for orientation under the influence of two beams of light, on the basis of a single assumption derived from experience with a single source. The fact that quantitative theory can here be tested by experiment only under certain simple conditions is due not to a failure of the analysis but to mathematical intricacies attending the orientation of a non-living machine. The problem of orientation with *three* sources of light is, for example, practically unsolvable, merely as an exercise in mathematics. It is fair to point out, and necessary to do so, that the execution of the pertinent experiments requires recognition of variables which are often disregarded by those who most delight in pointing to the complexity of organic behavior. In fact, it is legitimate to say that the kind of vagueness often apparent in reports of investigations in this field seems to result not infrequently from failure to conceive clearly the possible end of the analysis in view.

Even when in a given compound field of illumination it is possible to account precisely for the amount of orientation exhibited by a phototropic organism, certain constants remain in the equations. It is tempting to regard these as in part "threshold" constants, and as such perhaps relating to the mechanism of central nervous adjustments between competing excitations. In an organism exposed to two sources of stimulation the resultant movements may be determined by the summation or by the balance of the respective excitations; with other suitable conditions the movement may be decided by one of the two excitations to the complete exclusion of the effects of the other. Such a "decision" may be taken to represent an elementary act of central nervous adjustment, and, given rational units in which to express the effects of the two sources of excitation when acting singly, its nature may be investigated by measuring its quantitative modification under controlled conditions. There should be procurable, in this way, a means of approach to central nervous functions which might hope to effect a rationalization of their nature (48-56, 73, 78). In this respect the quantitative description of tropistic orientation can become the kind of analytic tool earlier hinted at.

The primary difficulty in such studies is, of course, the essential requirement that the integrity of the organism shall not be violated. The very thing which we desire to learn about depends upon the absence of destructive interference; we must not employ such treatments as may provoke irreversible changes. There remain, free from such objection, certain additional definite avenues of approach. Taken singly and in combination, we have a right to expect that these may assist in the way desired. There is first the modification of the temperature of the organism, and the consideration of the relations between temperature and the frequencies or velocities of measurable activities. It happens that these relations "make sense" in terms of elementary physical theory, and the ways in which they can be modified experimentally (77) already yield suggestive and significant conclusions (47, 49, 61, 92). The process of excitation by light (85-91) permits similar treatment. The methods of genetics, combined with such procedures, make possible to a degree the unification of the experimental objects as uniform and fairly comparable individuals, and in a few cases (where enough is known of the genetic behavior) even permit the making-up, at will, of individuals possessing particular characteristics required for specific purposes. At the moment there is to be stressed merely the importance of "analysis without interference."

IV

The quantitative description of photic orientation under conditions which exclude mechanical interference can, however, be carried out with some success (55, 111). Consider a negatively phototropic animal on a horizontal surface illuminated by two point-sources of light (Figure 11). The X-axis of a system of reference coordinates passes through the lights, the Y-axis bisects the distance between them. The animal at any moment has its anterior end A at a point (x, y), and the coordinates of the lights L' and L'' are respectively $(-a, o)$ and (a, o). Then the distances AL' and AL'' are $m^2 = (x+a)^2 + y^2$ and $n^2 = (x-a)^2 + y^2$. The dimensions of the photoreceptors are safely assumed small in relation to m and n. The lights L' and L'' have luminous intensities P' and P'' candles respectively. Orientation should be accomplished and maintained when the photochemical effects on the two photoreceptive surfaces are equal. This is the *only* assumption of the tropism doctrine, and it turns out to be sufficient. Taking first the case in which the organism is attached to the substratum, but able to orient, the position of orientation is defined by the positive angle θ which its axis makes with the X-axis connecting the lights, and it may be shown that the necessary relation between θ and the other quantities concerned is:

$$\tan \theta = \frac{P'n^3\,[\tan h\,(a+x)+y] + P''\,m^3\,[\tan h\,(a-x) + y]}{P'n^3\,[\,a+x-y \tan h] - P''m^3\,[a-x-y \tan h]},$$

where h = the angle between a photoreceptive surface and the longitudinal axis of the body. For an ideal organism which is *positively* phototropic similar equations are obtainable which differ only in detail. The formula given and simpler expressions of it easily deduced for various cases hold equally well whether the primary photosensory effect is proportional to the intensity of the light I or to log I. With negatively

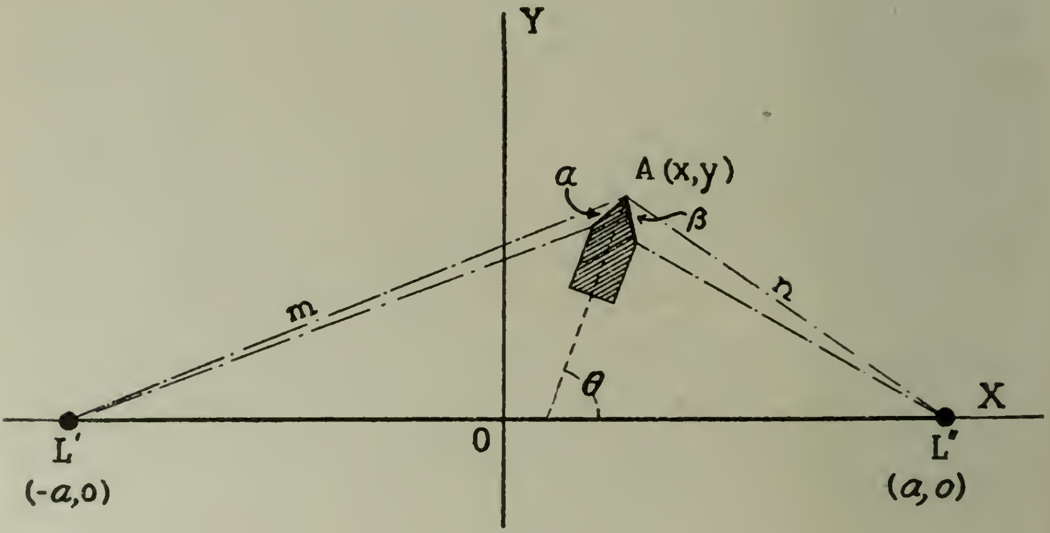


FIGURE 11

The Photonegative Organism at A (x, y) Oriented by Point-Sources of Light L′ $(-a, o)$ and L'' (a, o)

(From *J. Gen. Physiol.*)

phototropic organisms illuminated by beams of parallel rays opposed at 180°, the equation reduces to

$$\tan \theta_1 = \tan h \frac{I'+I''}{I'-I''};$$

with beams crossing at an angle of 90° (parallel rays),

$$\tan \theta_2 = \frac{(I'+I'')\ (1+\tan h)}{(I'-I'')\ (1-\tan h)};$$

[cf. (55), where these equations are directly deduced, but with $H/2 = h$, $\theta = 90° - \theta_1$ for the first case, $\theta = \theta_2 - 45°$ for the second]. When the light rays are parallel, there is no necessity for restricting forward movement in testing the equations.

These equations can be tested by a number of series of measurements. In some forms, as the larva of the blow-fly, the angle h is constant. With other forms it is zero, and in such forms the fact that $h = 0$ can be directly checked by failure to obtain orientation with lights opposed at 180°. A further check is given by the known facts of orientation (e.g., in experiments due to Buder, 1917-19) for forms in which $h = 0$, namely, various protista swimming in a helical ("spiral") path, and in which beams of light are used which cross at an oblique angle. In this case, from the general formula one obtains $\tan \phi = \frac{\sin \psi}{I'/I''+\cos \psi}$, where ϕ is the angle of the path of oriented movement with the axis of one of the two beams, and ψ is the angle between the beams. The agreement with experiment is extremely good.

The angle h is not constant, however, in all cases. In the blow-fly larva the sideward swings of the anterior end are connected with the mechanism of progression, and their frequency and extent are independent of illumination. With certain other animals this is not so. If photokinetic excitation results in "random" movements of the head, the frequency, speed, and amplitude of such movements should influence the calculated value of h, and this should be a function of the total intensity of excitation; h would be expected to decrease in proportion either to (I_1+I_2) or to $\log I_1 I_2$. In young rats (before the eyelids have parted) this is its behavior (Figures 12, 13) (55, 72). In other forms h increases with low total intensities of illumination, due to the suppression of random motions resulting from other kinds of excitation, and then passes through a maximum as the highest intensities provoke increasing "random" movements of the head. It is to be noted that there is obtainable in this way a *measure* of "random" movements and a key to their interpretation.

When these considerations are applied to the case of a moving animal in a field illuminated by two point-sources of light in the same horizontal plane, further difficulties arise. To distinguish this situation from that already treated we may speak of orientation of a *sessile* form as presenting

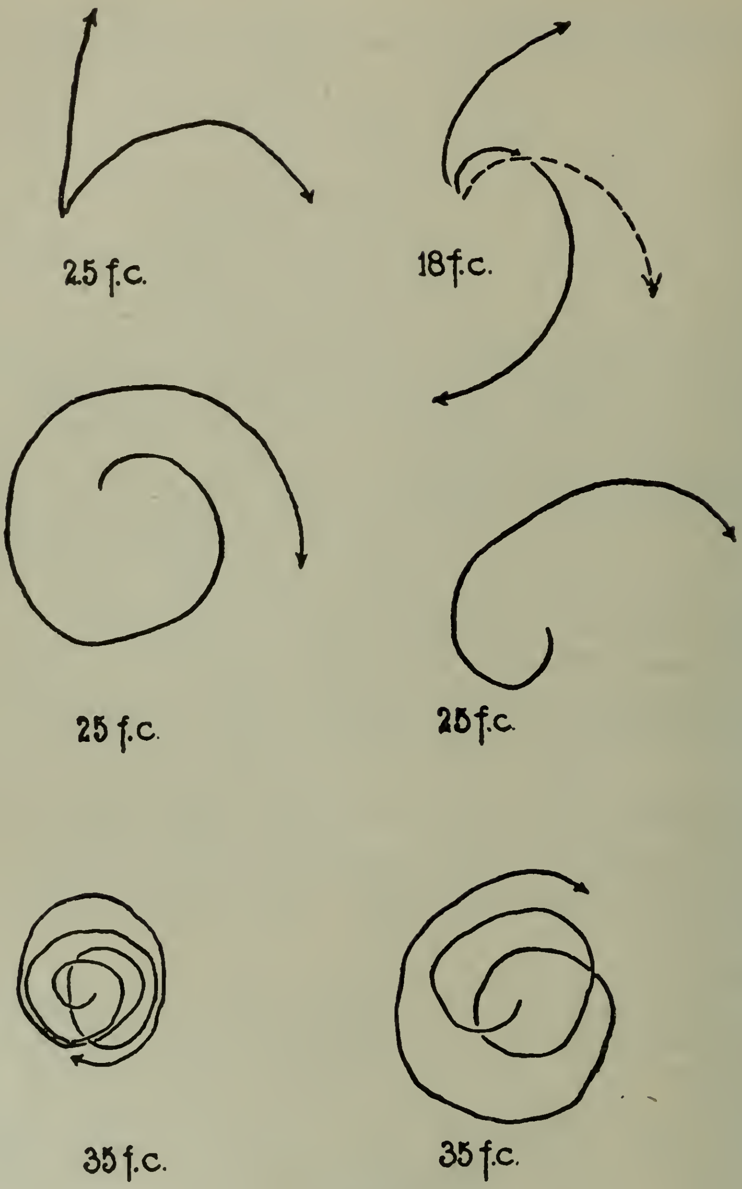


FIGURE 12

Circus Movement Trails of Young Rats with One Eye (the Right) Removed, the Eyelids of the Other Not Yet Parted

Illumination from a single source, vertically above the center of the creeping stage. Intensities of the lights (foot candles) measured photometrically at the creeping stage. Scale 1:10. These trails are typical. The amplitude of the orientation movement increases with the intensity of the light This is in part due to the fact that the animal creeps little but turns constantly. The records were obtained by following the positions of a marked spot at shoulder level on the animal's back, as seen against a system of coordinates in white lines upon a black cloth covering the creeping stage. (From *J. Gen. Physiol.*)

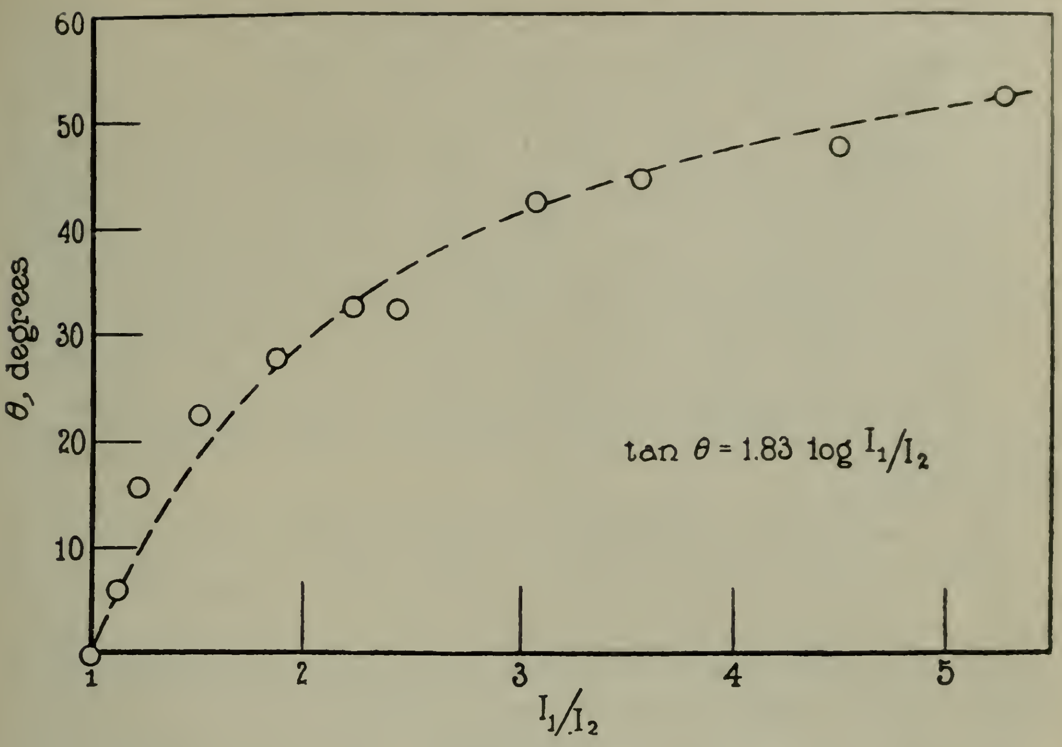


FIGURE 13*a*

The empirical formula $\tan \Theta = K \log I_1 / I_2$ gives an approximate description of the variation in angle of orientation as dependent upon I_1 and I_2, and may be used in order to obtain a graphical representation of the results.

(From *J. Gen. Physiol.*)

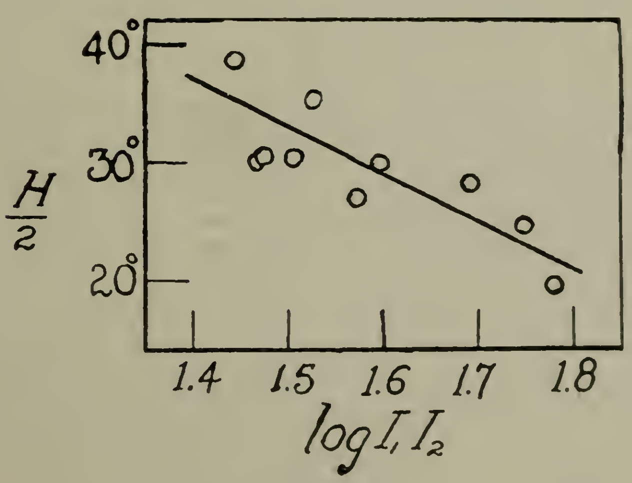


FIGURE 13*b*

VALUES OF $H/2$ COMPUTED FROM EXPERIMENTS WITH YOUNG RATS

$H/2 = -K \log I_1 I_2 + C$ (see text). (From *J. Gen. Physiol.*)

a problem of orientation *in situ;* the orientation of a motile form in beams of parallel rays is, of course, dynamically identical with this; the phototropic progression of a motile organism may be termed orientation *in transitu* (11). Here, the path of the animal from any point in the illuminated field depends upon the direction in which it is moving when subjected to the lights at that point. Differential photochemical effects in the two photoreceptors affect the rate at which the slope of the path varies with time (or with respect to distance covered, if the velocity of creeping be sensibly constant). The change of slope is $d\theta/dt$, but we cannot at once decide whether in such an instance the rate of turning depends upon the *difference* between the excitations on the two sides, or upon their *ratio;* and in either case there may be a direct or a logarithmic relation between excitation and the luminous flux. If the rate of creeping is very small, we have (from the general formula already given):

$$\frac{dy}{dx} = \frac{P'n^3\,[(a+x)\tan h + y] + P''m^3\,[(a-x)\tan h - y]}{P'n^3\,[a+x-y\tan h] - P''m^3\,[a-x-y\tan h]}\,.$$

Since m and n are complicated functions of x, y, and a, the expression is practically unsolvable as a differential equation, although it can be tested by several modes of approximation (cf. Figure 13*c*).

The elementary conditions for movement oriented by light have been reviewed in order to bring out the basis for a particular conclusion. Regarded simply as the orientation of a machine with bilaterally disposed photoreceptors (the symmetry may be dynamical, as in helical-swimming protista, rather than merely of an obvious structural sort), the description of phototropic movement is seen to present a very fair degree of com-

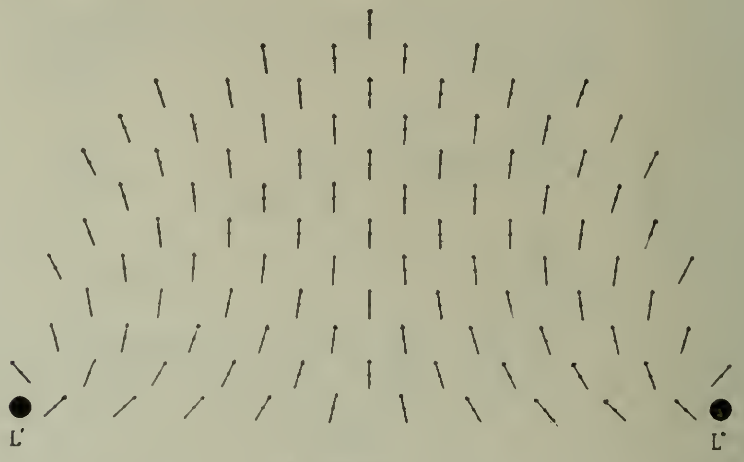


FIGURE 13*c*

Arrows Show the Values of θ for a Photo-Negative Organism at Different Points in the Field

The point-sources of light L' and L'' are equal in luminous intensity and $\tan h = 0.9$. (From *J. Gen. Physiol.*)

plexity. This becomes apparent at once if we seek to predict the character and extent of orientation when more than one source of light is acting, on the basis of what is known concerning phototropism in the simplest kinds of situations. Inability to deal quantitatively with the measurable aspects of the orientation of organisms in complex fields of illumination, therefore, is not to be taken gratuitously as evidence either of incomplete understanding of the case, nor, on the other hand, as indicating supra-mechanical influence at work within the organism. In an elementary way the kinds of complexity revealed by the present discussion serve to demonstrate concretely the justification for quantitative investigation under the most rigidly simplified conditions attainable, to which testable formulations can be applied.

It is not always true that the simplest conditions will be found in the most obvious place. An especially clear instance is found among galvanotropic phenomena. In the case of annelids (114) and certain other metazoans, the interpretation of galvanotropic orientation turns upon the necessary and sufficient assumption that the current serves to excite definite groups of nerve-cell bodies. This is consistent with the mode of action of strychnine, which produces a relatively specific reversal of inhibition in responses to peripheral excitation, and also reverses the mode of galvanotropic response. These effects seem to have considerable generality (54, 82). It is historically and in other ways of interest that galvanotropic movements were early reported as occurring in the roots of plants and subsequently examined in various animals. It turns out (120) that the "galvanotropic" curvatures of roots, which because of their slowness and the definite manner of their exhibition might be useful for quantitative investigation of the mechanism of stimulation in a structurally simple system, are, however, artifacts due to chemical products of polarization at the electrodes ordinarily used: the "galvanotropism" disappears if the electrodes are made non-polarizable.

V

The compounding of tropistic modes of behavior has been relatively little studied in any exact way, partly because the necessary preliminary descriptions have not been available. The balancing of phototropic and stereotropic excitations, for instances in which it is possible to show that the stereotropic effect obeys the laws of compounding already indicated for phototropism (67, 48, 49, 54), provides one illustration. Thus with *Tenebrio* larvae it is possible to obtain, from measurements of the intensities of lights of different wave-lengths required to counterbalance stereotropic adherence to a glass surface, a picture of the absorption of light by the photoreceptive organs. By compounding the effects of galvanic and photic fields mutually perpendicular, it is possible (78) to find for each of several intensities of light a current density which forces orientation of marine planarians (*Leptoplana*) at exactly 45° to the axis of either field. The galvanic excitation must then be equal, in its differential tonus effect, to the photic. Since the latter is practically pro-

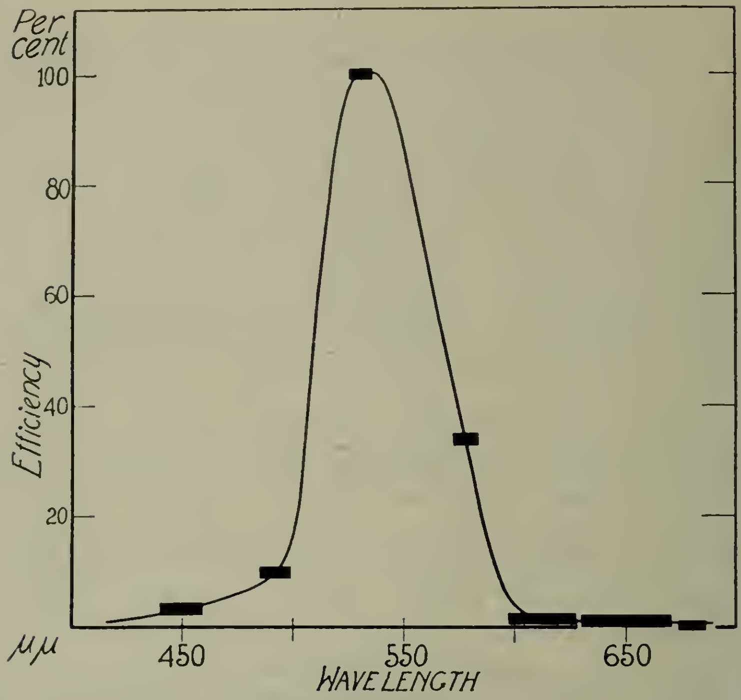


FIGURE 14

Measurements were made of the intensity of lights of different wave-lengths (obtained by filters of known transmissions) required to force creeping *Tenebrio* larvae to orient away from a vertical glass plate with which they otherwise maintained stereotropic contact. The light passed normally through the plate. Reciprocals of these intensities (i.e., the efficiencies,) measure the proportionate absorption by the photoreceptive system at these wave-lengths. The maximum of effectiveness is sharply localized. (Each observation is given as a bar, of which the width includes the zone comprising the wave-lengths embracing two-thirds of the energy transmission of each filter.)

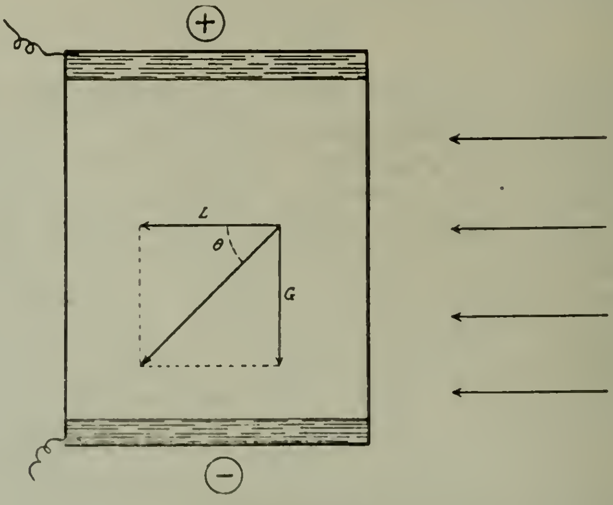


FIGURE 15

Diagram of Experimental Trough with Cotton Lead-In Pads (Lined), for Establishment of Constant Current Giving Rise to the Galvanotropic Vector, *G*

Light from the right side (arrows) is responsible for the phototropic vector, *L*. The galvanotropic excitation is equivalent to the phototropic when the resultant path makes an angle $\theta = 45$ degrees. (From *J. Gen. Physiol.*)

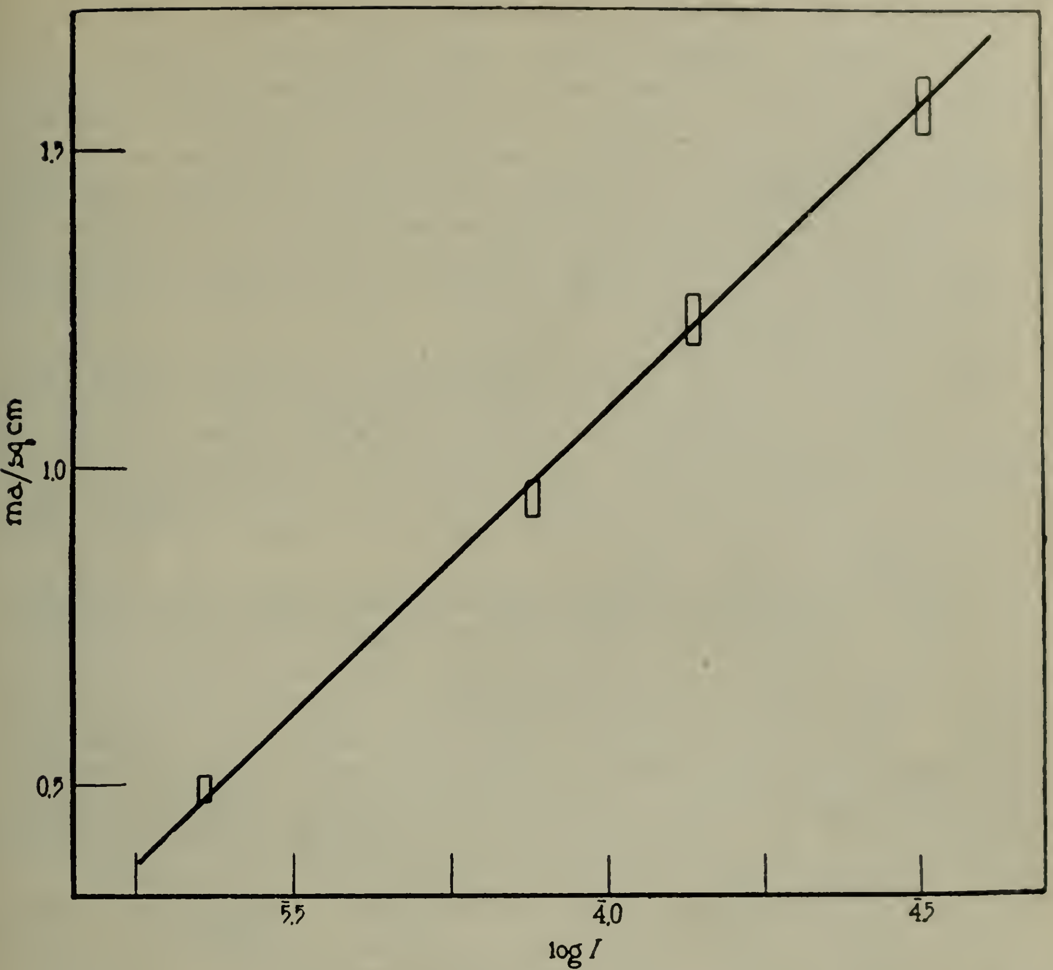


FIGURE 16

The current density in milliamperes per sq. cm. is directly proportional to logarithm of the relative intensity of the light, for orientation at 45° to the path under either mode of stimulation alone. The observations are plotted as bars centered on the mean current densities and having a vertical height equal to twice the probable error. (From *J. Gen. Physiol.*)

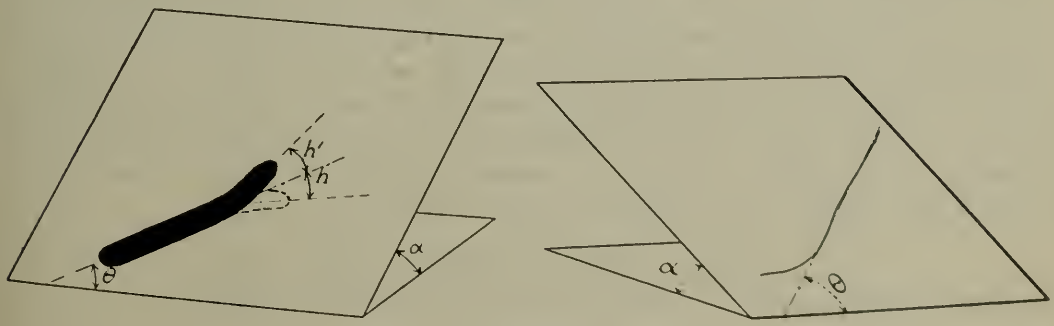


FIGURE 17

Illustrating the Definition of the Angles α and θ Involved in the Description of Geotropic Orientation of *Malacosoma* Larvae

The traced line of progression of the caterpillar, when creeping is started with the axis of the body horizontal, comes to make a definite angle θ on the inclined plane. In many instances the path is more complicated. (From *J. Gen. Physiol.*)

portional to log I, we expect the current density (for 45° orientation) to be a rectilinear function of log I. This turns out to be true. The organism behaves, in this case, not as if the compound situation involves properties peculiarly its own but as it should if responding to each kind of excitation separately; the effect is purely additive. Moreover, one secures a measure of the galvanic excitation in energy units even without knowing how much current goes through the organism; and since the excitation is found proportional to the current density, one has direct proof that the current probably stimulates by moving ions—as is reasonably assumed but without direct demonstration.

This procedure may be pursued further by following the consequences of compounding the influences of photic and gravitational effects. To do so requires a formulation of gravitationally controlled orientations, which may incidentally be used to illustrate means whereby the significance of descriptive equations may be tested. We shall confine attention to geotropic movements carried out upon a substratum, neglecting the problem of orientation in space (16).

A negatively geotropic animal, such as a tent-caterpillar, orients upward upon an inclined plane, but, in the absence of interfering modes of stimulation (i.e., under non-stimulating weak red light), it does not creep "straight upward" unless the surface upon which it moves is nearly vertical. The upward orientation ceases at an angle θ with the intersection between the creeping plane and the horizontal. At various inclinations of the surface (a) it is found that the average value of θ (or the geometric mean) is a definite function of a, such that, purely as an empirical relationship which has useful properties, $\Delta\theta = K\Delta \log \sin a$.

If we seek to obtain the reason for this effect, the conditions of progression must be examined with care (Crozier and Stier). It is to be noticed that the anterior end (head and thorax) of the caterpillar swings from side to side during locomotion. The possibility of a control of the angle θ is therefore given if we assume that orientation ceases when the difference between the pull of the weight of the anterior end of the body (freed from the substratum at the execution of each lateral swing) in the "up" position and in the "down" position is reduced to a threshold magnitude. Since the active component of gravity, in the plane of movement, is obviously proportional to sin a, this assumption is expressed by the equation $K \sin a[\cos(\theta - h) - \cos(\theta + h')] = -\text{const.}$, where h and h' are the mean angles between the axis of the anterior swung portion of the body with the main axis, in the "up" and the "down" positions respectively. There are means of showing that h' and h are statistically equal. The equation at once reduces to

$$(\sin a)\quad(\sin h)\quad(\sin\theta) = -K'.$$

If h be sensibly constant, and independent of θ (in the average),

$$(\sin a)\quad(\sin\theta) = -K''.$$

If a be constant, θ must vary inversely with h. By means of trails upon smoked paper, permitting measurement of the lateral excursions of the

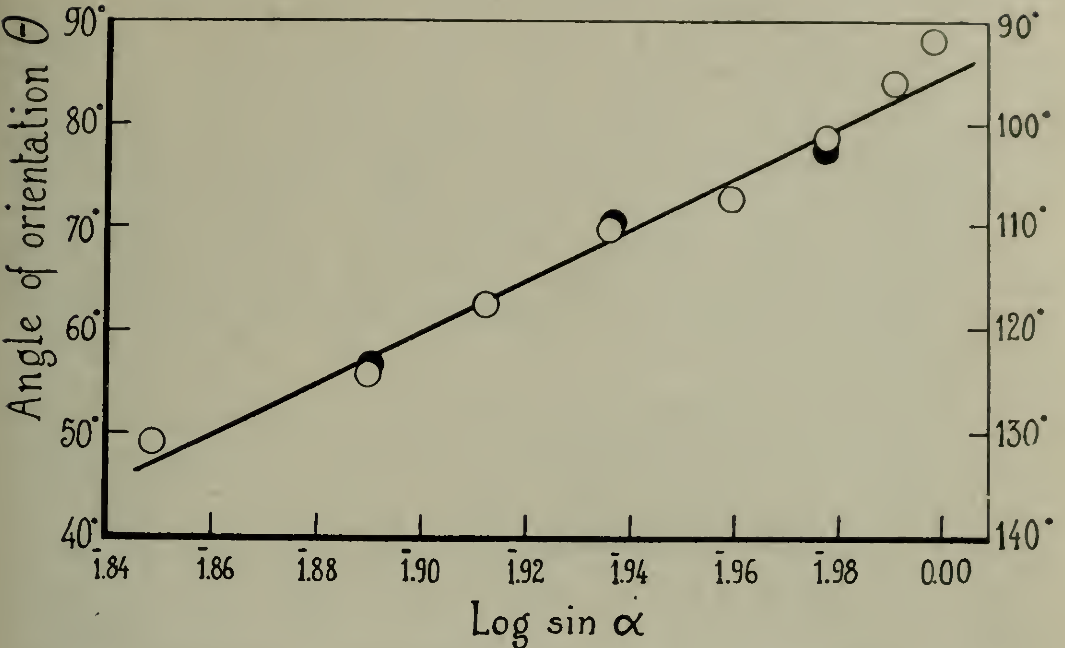


FIGURE 18

The angle of orientation (Θ) is directly proportional to log sin a, where a is the angle of inclination of the creeping plane. The solid circles are from measurements made at inclinations greater than 90° (i.e., the animal hanging from the under surface of the plane).

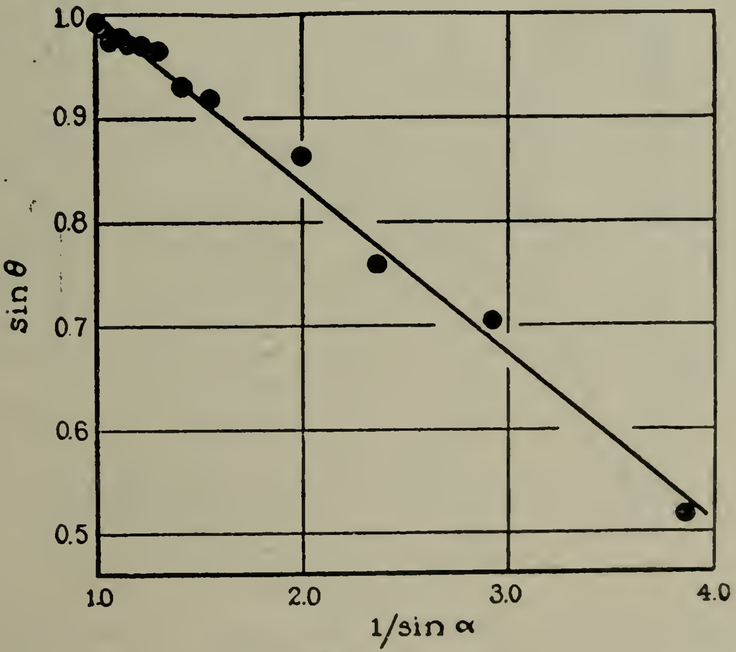


FIGURE 19

$\sin \Theta = -K/\sin a + C$

(From *J. Gen. Physiol.*)

head, the latter assumption can be tested and is easily verified. When a large number of trails are studied (about 3,000) this variable cancels out, and we should find sin θ inversely proportional to sin α. This expectation is thoroughly realized in the experimental results. This case is of interest in several ways. The "head angle" is similar to that previously involved in the discussion of phototropism, and is essentially a dynamic thing rather than directly a matter of structural symmetry. These facts are to be compared with the geotropic orientation of such an organism as the slug *Agriolimax* (140), where a *structural* angle of the required sort is readily perceptible; in this case also the formula just given is definitely obeyed. As an empirical relationship, it is in addition found that the mean orientation angle (θ) is very nearly a straight-line function of log sin α.

It is to be recognized that such formulae describe *positions* of orientation, and only indirectly have to do with questions of the *intensity* of stimulation. This should be sufficient in itself to prevent one from regarding the logarithmic formulation as an example of the Weber-Fechner rule. While it is impossible to convert the logarithmic formula into the second, trigonometric one, as a matter simply of mathematics, the former retains a definite utility in connections in which the trigonometric expression is unmanageable. Moreover, as will be indicated, it is possible to obtain some insight into its real meaning.

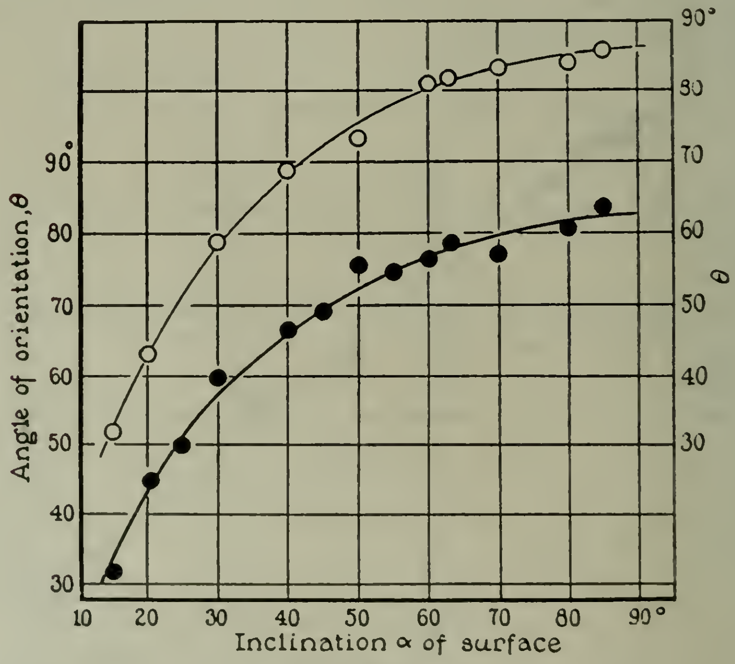


FIGURE 20

Relation between Tilt of Angle of Inclination (α) of Plane from the Horizontal and the Mean Angle of Upward Orientation (Θ) on the Plane (*Malacosoma* Larvae)

Lower curve, means of the average Θ's; upper curve, means from one individual. The curve is that for $\Delta\Theta/\Delta \log \sin \alpha =$ const. (From *J. Gen. Physiol.*)

The first experiments looking toward this type of formulation were made with young mammals (70, 71, 73, 75, 76, 128, 69, 98). Such experiments with rats and mice require special attention to the state of the test animals, their care and handling, and to the conditions of the tests. Since the results are susceptible of a particular kind of mathematical treatment, it is important to realize that the results of this treatment can be checked in several independent ways. Since the conditions of creeping in young rats change rapidly with age during the time within which it is possible to make tests having the desired significance, it is necessary to average results from a number of individuals. The point of the experiment consists in determining the relationship between the angle θ, already defined, and the inclination of the surface (α). To do this, individual litters are selected from lines of rats long inbred by brother-sister or other close mating. This insures genetic uniformity of individuals, and can be demonstrated to be a necessary precaution. Small litters are chosen so that all the rats may be vigorous creepers; this also is necessary. At the age of 12 to 14 days, before the eyes have opened, such individuals, with very rare temporary exceptions in the stocks we have employed, orient upward upon a surface tilted at 15° to 20° to the horizontal or greater. In a ventilated dark room where the temperature is 22° to 25°, observation is possible either under non-stimulating red light of low intensity or (in darkness) by means of markings with luminous paint. Accurate record is made of the paths pursued. At any one inclination the probable error of the mean θ is quite small, if due attention is given to the selection of the litters at the start, and to their intermittent replacement with the mothers for feeding. Just preceding the opening of the eyes the variability of behavior is definitely increased; this is seen also in mice. Low intensities of white light, before or after the opening of the eyes, render it very difficult to secure uniform orientations. This effect cannot be obviated by surgical removal of the eyes, which enhances irregularity of creeping and accentuates pauses; in these animals, as appears to be the general rule, geotropic orientation is evident only *during* progression. The consequence of removal of the eyes is undoubtedly comparable to that seen in certain fishes (45). The desired test can, however, be made in another way, using a gene which prevents the development of the retinal rods; mice carrying this gene are blind and can be used for the geotropism experiments in daylight (98).

As a sufficient approximation, θ is found proportional to log sin α. The graph is more exactly a long-drawn *"S,"* however, as every series of experiments has shown. The logarithmic fitting is sufficient, none the less, for certain purposes. More exactly, considering the way in which the weight of the body is carried by the legs, it is possible to show that cos θ would be expected to decline in direct proportion to sin α. This expectation is precisely obeyed, and the constants in the resulting equations are recoverable in new tests with rats of the same stock after an interval of over a year (76). The data from such experiments can be submitted to a variety of independent tests, which can be only very briefly referred to here. The *speed* of upward progression obeys the same rules as θ (128); from this fact, computations can be made as to the rate of doing work when the body is lifted, as a function of α. The effect upon θ of adding weights, attached to the root of the tail, which is predicted in the derivation of the relation $\Delta \cos \theta / \Delta \sin \alpha = -$ const., is also found experimentally (75, 128). The *variability* of θ at successive magnitudes of α shows that the measurements contain within themselves an automatic check upon their significance. If θ is determined by α in the manner indicated, increasing sin α should *decrease* the variability of θ according to the same rule as that by which θ itself increases. This is due to the fact that increasing geotropic excitation not only involves a greater amplitude of response, but proportionately cuts down the effects due to unavoidable sources of stimulation not directly concerned with the experiment. We find, in fact, that the statistical measure of the variability of θ decreases in direct proportion to log sin α. Moreover, if this conception be sound, the *rate* of this decrease should depend directly upon the number of individuals involved (the total number of observations of θ being the same in each case). A litter of two individuals indeed shows one-half the rate of decreasing C. V. θ as log sin α increases, as is shown by a comparable one of four individuals.

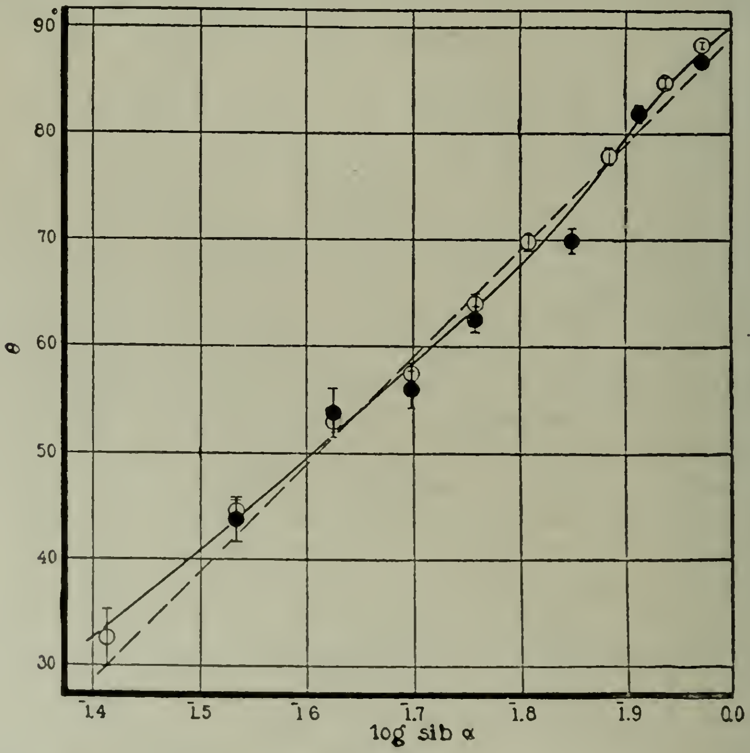


FIGURE 21

Showing the agreement in the relation between amplitude of negatively geotropic orientation Θ and log sin α, where α is the angle between plane of creeping and the horizontal, in two series of tests with young rats (*R. norvegicus,* line *K*) made 15 months apart. The height of the vertical bars = 2 P.E. For certain purposes the relation may be regarded as effectively rectilinear, though in reality sigmoid. (From *J. Gen. Physiol.*)

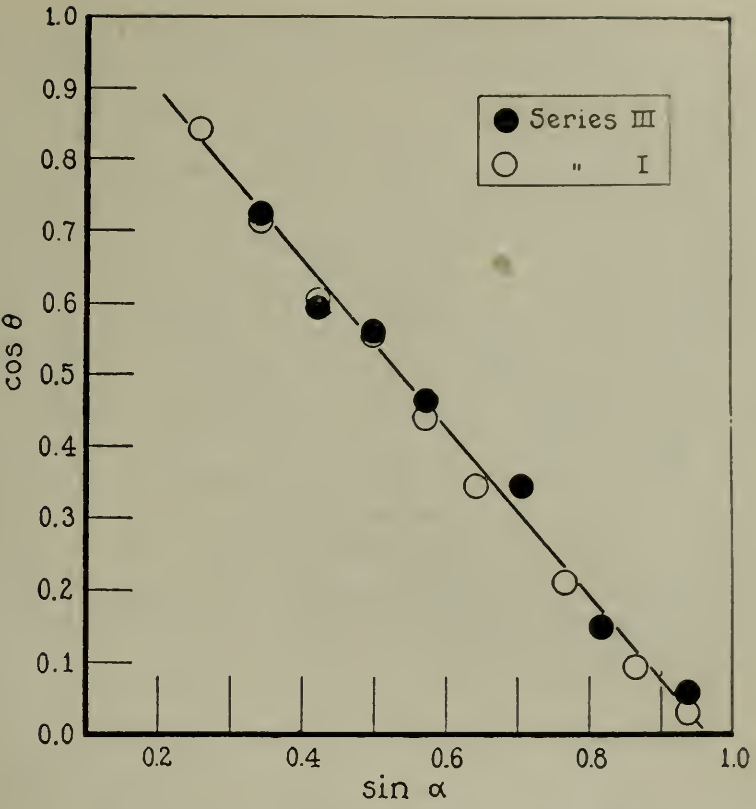


FIGURE 22

THE AGREEMENT OF TWO INDEPENDENT SERIES IN TERMS OF A LINEAR CONNECTION BETWEEN cos Θ AND sin *a*
(From *J. Gen. Physiol.*)

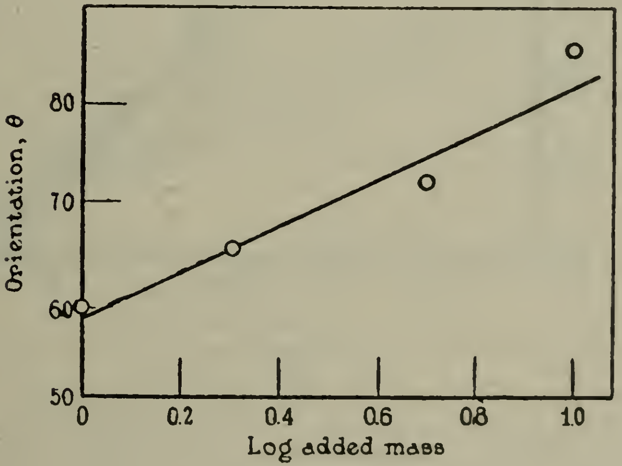


FIGURE 23

The angle of upward orientation is proportional to the logarithm of the mass added to the tail of the rat, when the inclination of the creeping plane is constant. (From *J. Gen. Physiol.*)

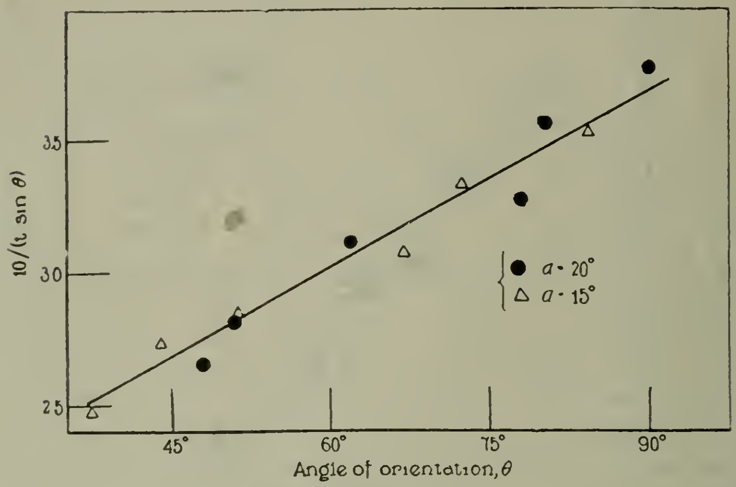


FIGURE 24

The rate of creeping with attached weights at two angles of inclination (15° and 20°) is plotted against the observed angles of orientation at these angles of inclination. (From *J. Gen. Physiol.*)

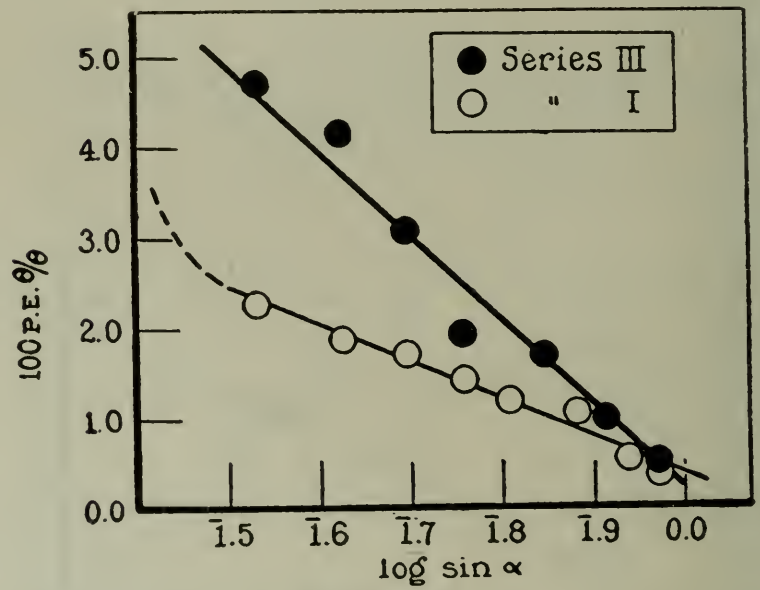


FIGURE 25

The variability of θ, expressed by P.E. θ as a percentage of the mean (the number of variates being 40 in each case) declines directly as log sin *a* increases. The rate of the decline is directly proportional to the number of individuals concerned: in Series I, 2 individuals, in Series III, 4, the slopes of the lines being in the ratio 1:2.1. (From *J. Gen. Physiol.*)

The mathematical formulations of geotropic orientation thus far presented are either such as may have a real meaning, or else they must be the happy results of a species of accident in some way due to the averaging of observations. If the latter were true, then it is scarcely conceivable that animals creeping in quite different ways, in which the possible manner of support of the body upon an inclined surface is completely different, should provide specifically different empirical relationships between θ and α, or obey diverse rational equations. The beetle *Tetraopes tetraopthalmus* has been studied in this connection. The body is in contact with the supporting surface at the tips of the legs and at the posterior end of the abdomen; θ is found proportional to sin α directly, *not* to log sin α. The complete formulation for this case is difficult, and need not here be considered further. Essentially, the magnitude of θ is determined by the fact that the leg muscles not only support the downward pull of the mass of the body, but are also subject to a torque due to the twisting movement of the abdomen, the center of gravity being posterior to the point of attachment of the legs. This interpretation is readily controlled by the attachment of additional loads (wax) to different parts of the beetle's body. When a mass of wax is attached to the abdomen, θ is increased. If attached to the anterior end of the body, θ is *decreased,*

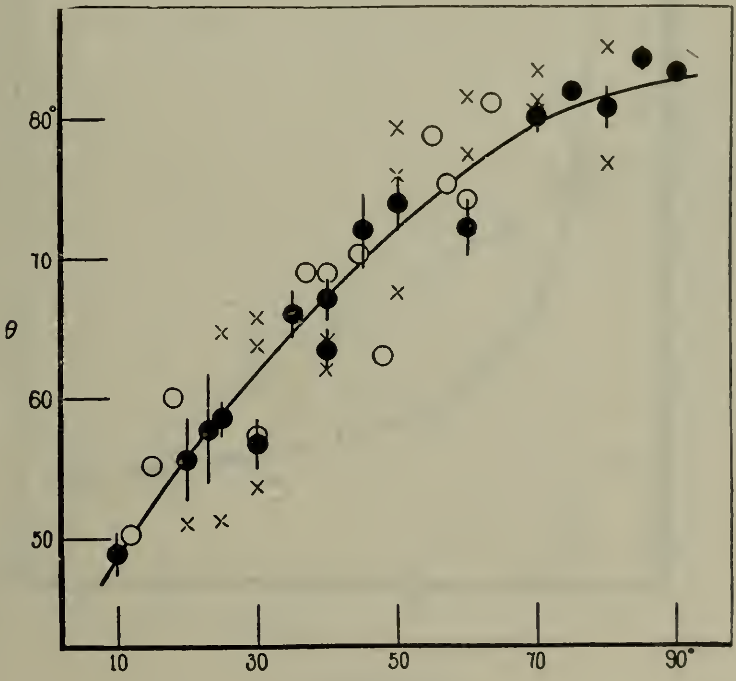


FIGURE 26

Mean magnitudes of angle of upward geotropic orientation of the beetle *Tetraopes tetraopthalmus,* as a function of the inclination of the surface α. Values which are averages from fewer than 10 series of measurements are given as open circles, of 10 or more as solid circles (with the probable error of the mean of the series' means indicated. There are added (crosses) mean estimates of Θ from tests with a single individual (No. 15) used at a large number of inclinations. The curve is that of $\Delta\Theta/\Delta$ sin α = const.

although the total load carried has been augmented, because the torque due to the abdomen is reduced. Again, with these insects, it is possible to perform the converse experiment and to *reduce* the load carried, and at the same time the torque on the legs, by amputating the abdomen. If our understanding of the situation is sound, θ should then be decreased, which is in fact the case.

These instances, and a number of others, can thus be understood as manifestations of geotropism which depend upon the distribution of the pull of the organism's mass upon the supporting musculature. The analysis does not exclude the involvement of statocyst function in geotropism, but leaves to be demonstrated just what this function may be in specific instances (28, 29, 40, 64). Related to this matter is the failure to obtain clean-cut relationships between θ and a in the case of certain slowly moving marine and fresh water gastropods; but this point cannot be developed at present. I must also refrain from going into the very attractive questions arising when geotropic orientation is reversed. Two groups of facts should be mentioned, however, as concerned in the demon-

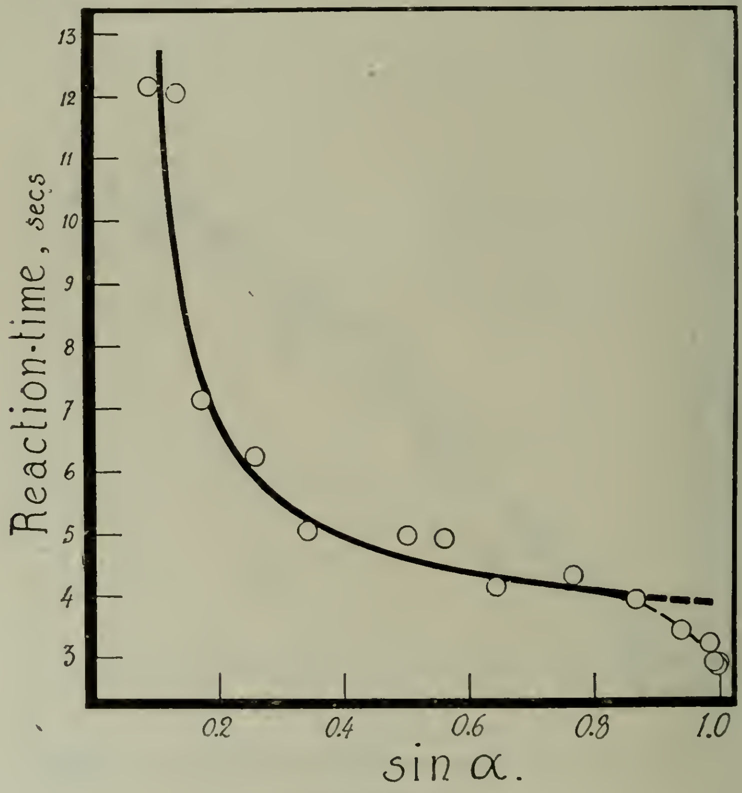


FIGURE 27

A snail (*Helix*) creeps straight upward upon a surface inclined at a° to the horizontal. When this plate is rotated smoothly but quickly in its own plane, the snail orients to a new upwardly directed path. The reaction-time, measured from the point of mid-rotation of the plane, decreases systematically as sin a increases. The latent period for geotropic response is thus a function of the magnitude of the exciting component of gravity. (Experiments by Hoagland and Crozier.)

stration that the pull of the mass of the body of a snail or slug upon its parietal and columellar muscles is a determining factor in geotropic response. When a snail such as *Helix* is creeping straight upward ($\theta = 90°$) on a sloping surface which can be rotated in its own plane, rotation through 90° is followed by orientation of the snail after a definite latent period. This latent period is a function of the active gravitational component, such that reciprocal of (latent period minus a constant) is directly proportional to the sine of the angle of slope of the surface; the velocity of the effects underlying gravitational excitation is thus determined by the gravitational pull. On a vertical surface the latent period for reaction, after 90° displacement of the axis of a creeping snail such as *Ligus* is systematically decreased by the addition of weights, is increased if the experiment is made under water, and decreases again (under water) if additional weights are carried (68).

The orientation of fiddler crabs may be taken as a further illustration of the fact that the empirical and the rational formulae applicable for mechanically diverse instances of geotropic orientation cannot be regarded

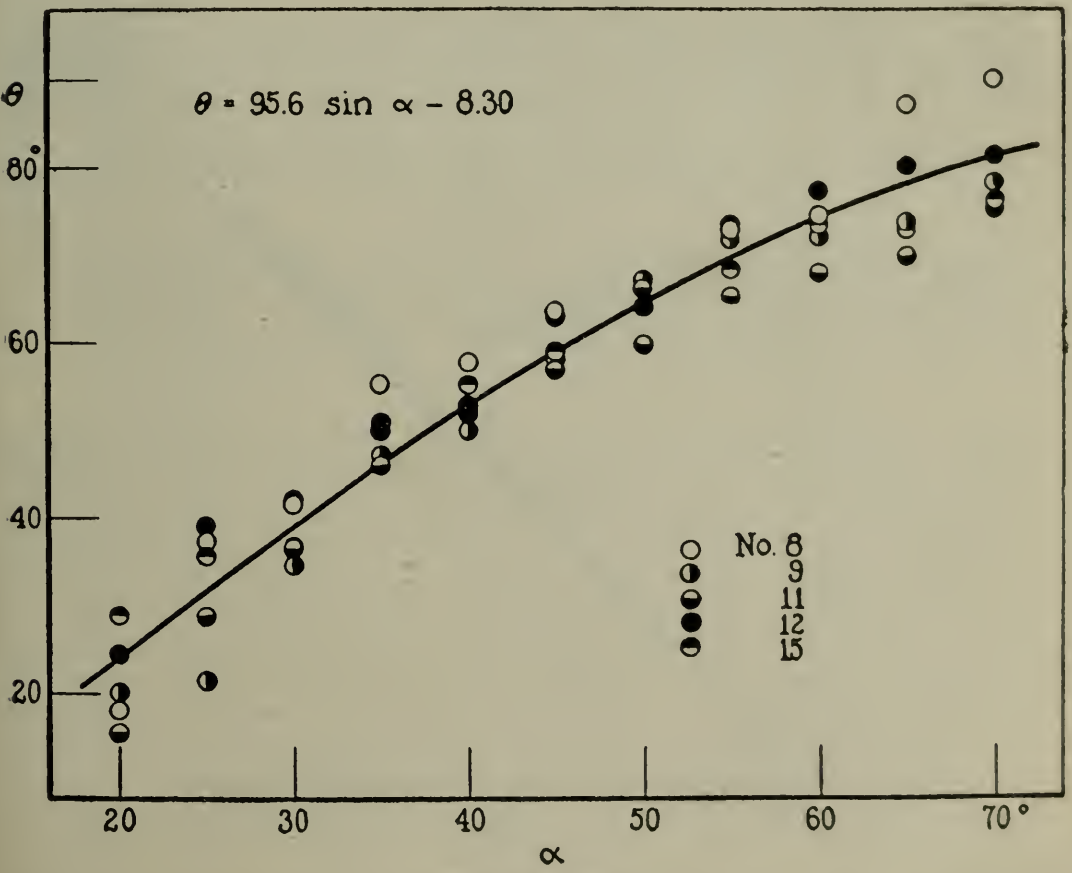


FIGURE 28*a*

MEAN MAGNITUDES OF Θ, THE ANGLE OF THE GEOTROPICALLY ORIENTED PATH OF PROGRESSION, FOR FEMALE FIDDLER CRABS ON A SURFACE WITH SLOPE α°

Each point is the average of 10 readings. (From *J. Gen. Physiol.*)

as "accidents," and that, in consequence, the muscle-tension theory of the limitation of orientation is not merely adequate but necessary. *Uca* creeps not straight forward but sideways. In the dark, Kropp and Crozier (99) found θ very closely proportional to sin α directly. If this is due to the attainment of mechanical stability, and thus to the release of a certain class of tensions on the legs, θ should be altered if the center of gravity is shifted. It happens that nature has performed the experiment. The large claw of the male fiddler may be carried "up" or "down" as the animal creeps sideways on an inclined plane. In the former case, θ should be less at each value of α, and this is the fact. Removal of the claw abolishes the asymmetry, which does not appear in the symmetrical females.

There is mentioned subsequently a group of tests with young guinea pigs in which the angles of stable progression were measured on sloping surfaces. The mechanics of movement here differs from that seen in the young rats, because the body is supported free from the surface. Again we should expect a different formulation to be required for the connection between θ and α, and in fact we do not find θ a (nearly) straight-line function of the slope, but that to a very good approximation log $\theta = K\alpha +$ const.!

Two facts have been referred to which it is necessary to bring into conjunction in dealing with a final aspect of this geotropic orientation. One is the fact that in the case of rats and mice the graph of θ *vs.*

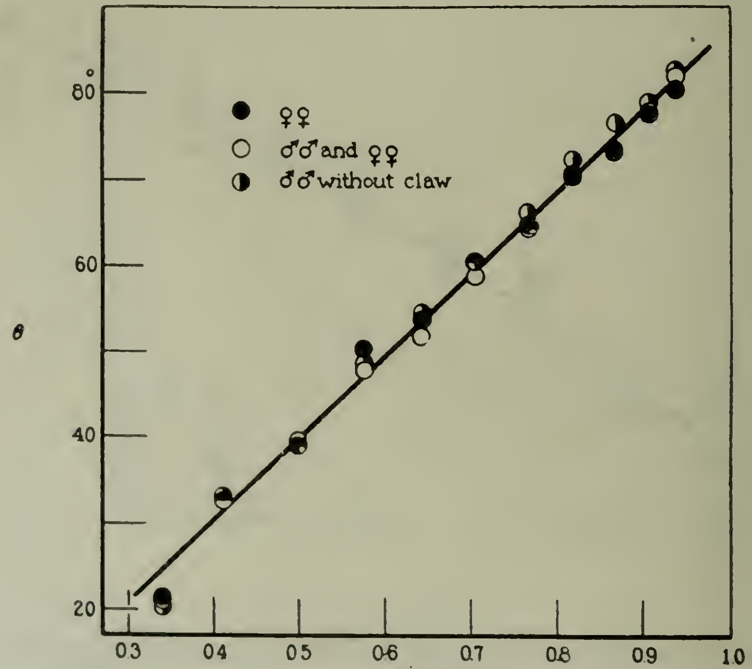


FIGURE 28*b*

Mean Values of Θ for ♀♀ and for ♀♀ and ♂♂, Obtained as Described in the Text, Give a Rectilinear Function with sin α

The same relationship holds for ♂♂ deprived of the large chela. The probable errors of the plotted points are less than 1° of arc. (From *J. Gen. Physiol.*)

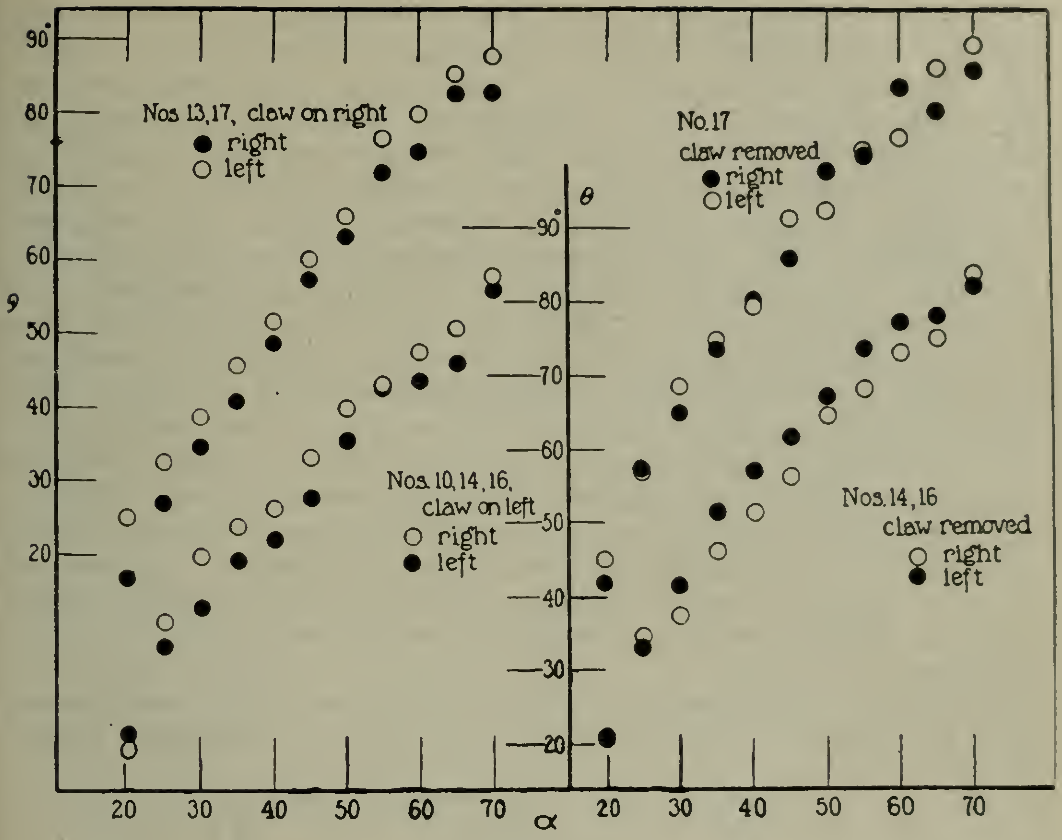


FIGURE 29

Mean orientation angles to right and to left for 2♂♂ *Uca* (13, 17) with large claw on the right side, and for 3 (10, 14, 16) with the large claw on the left, demonstrating lower Θ's when creeping is toward the side carrying the large chela. This relationship is obliterated when the large claw is removed. (From *J. Gen. Physiol.*)

log sin a is *almost* a straight line, but, in reality, a long-drawn *S*. The other is the necessity for genetic uniformity in the test material. This is proved by the fact that whereas the same values of the constants in the equations are repeatedly obtained with individuals from the same stock, it is nevertheless an easy matter to obtain separate stocks, of known different genetic history, and distinct species, in which (1) the slope of the line connecting cos θ with sin a is different, or (2) in which the slope is sensibly identical but the intercepts on the axes are different. Consistent performance within an inbred line of animals must be due to homogeneity of inheritance, though it does not follow that it should be easy to work out its actual basis. Yet the experimentation undertaken on this basis has shown that it is possible to learn something about it, and at the same time to obtain a neat justification of the initial treatment of the data.

Behind such attempts there persists a general problem, which has been more or less ignored as a problem although often a source of unintelligent mirth to those impatient with mathematical formulation. It has to do with the question as to the degree with which the behavior of an organism can legitimately be given mathematical formulation in an equation. The consistency of observations can, of course, be weighted by statistical methods, but this is not primarily the point. In other cases the general mode of treatment, the kind of interpretation provided, can be tested by deduction of the forms of hitherto untested but contingent relationships, subsequently studied in a quantitative way. But this does not completely answer the requirement, because not infrequently a question remains as to the biological "reality" of the assumptions employed in the original formulations. This difficulty is in part an emotional one, and is often justified. Of the two types of formulation consciously employed, that which is purely statistical is unsatisfying, since the constants it provides are dimensionless and arbitrary; expressions derived from physical theory have the enormous advantage that the character of their contained constants immediately gives basis for concrete experimental tests. Without such tests these equations may also remain arbitrary. If the organic significance of the constants is the objective of inquiry, it becomes necessary to justify in some independent way the specific equation chosen to express the initial results, or at least the mode of representation adopted for them.

The proof of the physical adequacy of a given mathematical formulation for events, behavior, in an organism seems to necessitate demon-

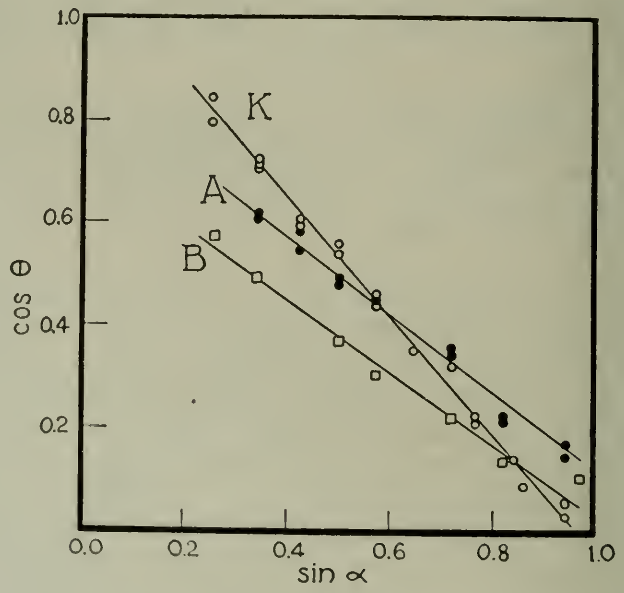


FIGURE 30

Data from several series of observations with young rats of race *K*, of race *A*, and of race *B* demonstrate the consistent differences with which the upward orientations (Θ) on a surface of slope α adhere to the formulation cos Θ = *a* − *b* sin α, where *a* and *b* are constants. (It will not do to fix the straight lines and determine the constants *a* and *b* for the purpose of obtaining Figure 31 by differentiation, because the departures of the observations from rectilinearity are systematic and significant, and it is these departures which in part determine the essential characteristics of Figure 31.)

stration that its constants have a "real" biological basis. It is easily objected that biological transactions are "too variable" to warrant such an undertaking, but there is ground for rejecting this view. It can also be objected, with greater cogency, that for any smooth relationship between two variables a large number of different formulations might be devised. This difficulty is easily avoided, however, if we confine attention to the attempt to show that the *difference* between two magnitudes of a dimensionally identical constant is biologically a real difference.

A "biologically real difference" would be one which, for example, behaves in a definite unitary way when subjected to manipulation of a purely biological sort. It must then be conceded a molecular substratum. A unitary effect in inheritance, exhibiting allelomorphic relationships with other similar units, is an example of such reality. For our further purposes, however, it is necessary to consider a "unitary effect in inheritance" as not precisely a *gene,* as the concept supporting this term is customarily used, but to consider it as a unit of inheritance having the dimensions of "amount of degree of organic expression *per unit of some* (*specified*) *controlling variable.*"

In our several genetically stabilized lines of young rats the extent of upward orientation on an inclined surface is a function of the slope of the surface. The form of this function is the same in the several races, but the numerical values of the respective constants are characteristic for each race. Figure 30 contains data from successive series of measurements with rats of lines K and A. Details of the procedure and of the statistical weights of the mean orientation angles are given in the original reports. In terms of the formulation previously given, $\Delta \cos \theta / \sin \alpha = - \text{const.}$, in each case, where $\theta =$ orientation angle on the surface inclined at α° to the horizontal. But the intercept constants and the slope factor are clearly different for the two lines. The graph for A is as if rotated about a mid-point of that for K. With an additional race, labelled B, we find that the slope $\Delta \cos \theta / \Delta \sin \alpha$ is sensibly the same as for race A, but that the threshold intercept is higher. There are thus two kinds of difference recognizable among the constants in the equation successfully describing the relation of amount of orientation to response.

It was earlier shown that to a fair approximation, the extent of upward turning is nearly a straight-line function of log sin α. But in fact every series of measurements shows that the relationship of θ to log sin α is to be described as a long-drawn S. (It happens that lines A, K, and B give straight-line relation between cos θ and sin α over the whole working range of sin α, a fact which has made the further analysis possible. In other genetically uniform groups of rats, different parts of the sin α range give lines of different slopes. The reason for this becomes obvious in the later treatment of the data.) The relation $\Delta \cos \theta / \Delta \sin \alpha = - \text{const.}$ was arrived at by considering that stable orientation on the inclined surface is achieved when the tension excitations due to the pull of the animal's weight are the same, within a threshold difference, on the legs of the two sides of the body. In this event the extent of upward turning is a direct proportional measure of the total amount of stimulation per (large) unit of time.

The total excitation of tension receptors must be regarded as involving, over a gross interval of time: (1) the total array of these receptors with thresholds below a certain value, which is a function of the stretching force; and (2) the frequency of change of tension. The latter, it is to be presumed, is determined chiefly by the frequency with which steps are taken, and should thus be proportional to the speed of progression. It has been shown that the speed of progression is a straight-line function of log sin α. Hence the differential, $\Delta\theta/\Delta$ log sin α, when plotted as a function of the stretching force, should give a picture of the distribution of effective thresholds among the available tension receptors in terms of the exciting component of gravity. For lines *A* and *K* these distributions are given in Figure 31. For these two lines the distributions are clearly resolvable in each case into three distinct groups. The only real difference between these curves has to do with magnitudes of the respective ordinates, which are in constant ratio throughout.

In the case of race *B* the three components of the array of receptors are again apparent, but the distribution curve for each one is broadened, the modes, however, occurring at the same points on the sin α scale.

It has been possible to show through tests in which a small mass of metal has been attached to the animal that these "groups of sense-organs" are selectively influenced by the added load, in such a way as to demonstrate the organic distinctnesss of the several groups.

Individuals of the F_1 generation produced by crossing lines *K* and *A*, and tested under the same conditions, show analogous relations between θ and log sin α (Figure 32*b*). Analysis of the differential of this curve demonstrates the presence of a lower segment identical with that for the *A* parent (parts *i* and *ii* of the curve in Figure 31), an upper segment entirely similar to that for the *K* parent.

A sufficient genetic test of the "reality" of the constants in the formulae descriptively connecting the response θ with the exciting component of gravity should thus be obtained by the study of individuals produced in the back-cross generations (F_1 x *A*) and (F_1 x *K*). In the first case the individuals should yield curves all of one kind except that in one half of them the terminal slopes of the curve of θ *vs.* log sin α should be steeper than with the other half. This is found to be true, although the differences sought are small. In the other back-cross, however, we look for four phenotypic groups of individuals, respectively showing: *I, II; i, II; I, ii; i, ii.* In the numbers of cases tested, these four classes are indeed clearly recognizable, by several independent criteria, and occur with equal frequencies. A summary of the data is given in Figure 32*c*.

We may accept the view that the definition of a *gene* should properly be based upon the formulation of an effect in inheritance as a function of some independent variable which can influence its expression. Otherwise no units can be obtained for the measurement of the gene. It may then be argued that in the present case the slope of the line connecting the amount of response with the inclination of the surface, that is, with

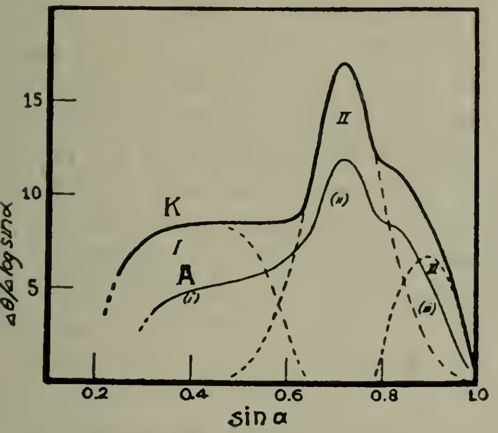


FIGURE 31

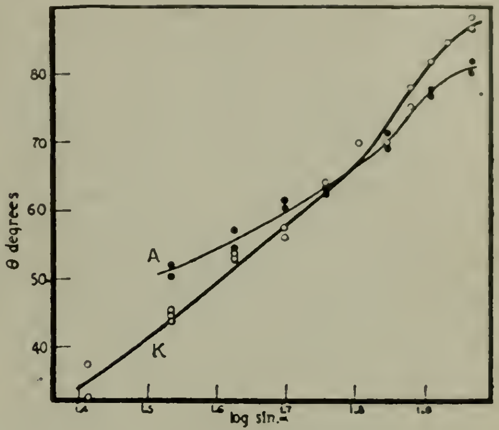
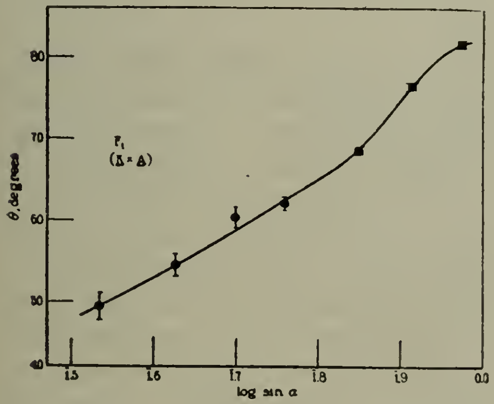


FIGURE 32*a*

Figure 31. The first differentials of the curves given in Figure 32*a* are plotted against sin α, to obtain a picture of the array of thresholds among tension receptors concerned in the geotropic orientation. The curves are clearly composite, and resolvable into three portions in each case.

Figure 32*a*. The amount of upward orientation, in degrees (Θ), is related to log sin α, where α is the slope, by a sigmoid curve; data from races *K* and *A*.



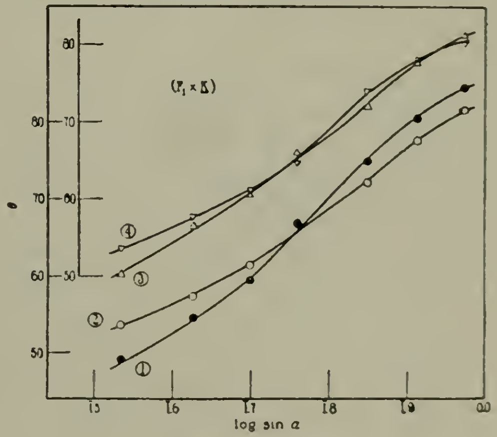


FIGURE 32*b*

FIGURE 32*c*

Figure 32*b*. θ vs. log sin α for the F_1 progeny ($K \times A$); points are plotted as bars with height = 2×P.E.

Figure 32*c*. θ vs. log sin α for the four groups of individuals segregated in the back-cross progeny ($F_1 \times K$). Group 1 should be like the grand-parental *K*; Group 3 should have its low-α end like that for 1, it mid-portion like that of Group 2. Group 2 should resemble closely, except at α = 55° — 70°, the grand-parental *A*, or more closely the Group (i. ii, III) of segregates in the back-cross ($F_1 \times A$), with which its curve is indeed identical. Group 4 should show a lower portion of its curve like that in the curve for Group 2, its later portion like that in the curve for Group 1. In other words, as regards the slopes of these curves, considering each curve divided into two parts on either side of α = 35°, we should find all possible combinations of low and high slopes—high slope below 35° combined with either high or low slope above 35° (up to α = 55°), and reciprocally. It is clear that these combinations are indeed found.

the exciting force, is definitely an example of the operative significance of this mode of defining a "unitary effect in inheritance." It is not the amount of the response which is subject to hereditary transmission but the way in which this response changes as a function of its inciting condition. Without such formulation, the real nature of the differences between lines *A, K, B* in the respects which here concern us could not be recognized, let alone analyzed. The fact that the case investigated is complex, in the sense that three pairs of "factors" have been distinguished, is in no degree adverse to this general statement, and indeed points the way to certain attractive theoretical possibilities. It should be mentioned, however, that the statistical significance of the differences relied upon for the analysis is fully protected through the demonstration that the variability of the responses is subject to exactly the same rules as those which connect the amount of response (θ) with the gravitationally exciting component ($\sin \alpha$).

The problem initially raised had to do with the possibility of demonstrating that the difference between numerical values of a dimensionally identical constant describing the homologous behavior of two different groups, genetically stabilized, of the same kind of organism, may have a "real" significance. The facts summarized show that the required demonstration seems to have been obtained, and that the specific values of the slope constants obtained in the examination of the races *K* and *A* are recoverable in a genetically rational scheme after two generations of cross-breeding.

This particular case, by itself, gives a specific illustration of the necessity for the preliminary establishment of biological uniformity among individuals concerned in quantitative experimentation. It happens that it does not stand alone, since facts of a comparable sort have emerged in the course of an analogous research involving temperature characteristics.

That the responses of an organism are not invariable to the degree witnessed in the activities of simple machines has often served as an excuse, if not indeed as an apology, for loose experimentation in relation to the conduct of animals, including that of man; and for obscure thinking in the discussion of experimentally obtained results. Two quite different modes of regarding this variability are of practical consequence. We may (1) consider that variation, "unpredictableness," is a category of phenomena exhibited by organic systems, presumably originating either as a result of the devastating complexity of such systems or as due to the participating interest of whatever god or devil one's private system of vitalistic "causation" may chance to be entertaining at the moment. Or (2) we may proceed in a quite different way, if we are sufficiently free from the urge to employ or to gather material for the support of some pet epistemological system—because we may know that data adequate for a "system" do not exist, and that any prevalent one we are likely to embrace is, after all, intellectually an impertinence as well as a handicap. It may be attempted to classify types and modes of "variability," because it is apparent that commonly quite a diversity of happenings have been loosely grouped under this label. Certain aspects of such classification have been referred to previously (62-64, 56, 70-76).

A thought which we may choose to guide us is that distinctions must be made between phenomena of mere *difference,* of *variation,* and of *variability,* if by the latter we understand "ability to vary." And as must always be the case, we may hope for light if we seek to express *variability* as a function of some significant measurable attribute of the system or situation which concerns us, and thus obtain a measure of "ability to vary." It is easily shown that, in spite of its very elementary appearance, the conception of *variability* based upon variation expressed as a function of some independent variable is, so far as one can detect, not usually employed. In the majority of cases, writers dealing with "variability of behavior" have been content to record differences in behavior; and to draw some obscure comfort from the fact that differences exist "under the same conditions." One excludes those multitudinous recorded instances in which some measurable feature of conduct has been shown to be *modifiable,* because the changes properly so described comprise a clearly separable assemblage of problems, and are not adequately cited as instances of variability; by the latter one intends to imply that the controllable conditions are the same, yet the responses vary. The distinction is well exemplified in connection with experiments upon photic excitation. With successive stimulations by constant increase of intensity of light, the speed of response in the case of such forms as *Ciona* (86) undergoes progressive orderly change; the modification of amplitude of response with *continuous* excitation by constant light is of an entirely comparable sort (141, 79, 80), and in each instance can be understood completely as due to sensory adjustment, in terms of a simple photochemical system which embraces also the phenomena of dark-adaptation (86, 87, 91, 42, 80). These are instances in which the energy change external to the organism, the "stimulus," or more correctly the *stimulating agency,* is obviously constant, yet the response of the organism is not the same; but its changes are not satisfactorily accounted for by mere reference to "variation," or "modification," because it is the precise form of the orderly change, as a function of time, which assumes significance. Without analysis, for which exact experimentation is necessary, in such cases as we have just received it would be easy to speak of "higher behavior," "learning," or similar effects. The possibilities become more elaborate when cyclic alterations of response are involved, but these we need not now discuss.

The point for which we are contending is sufficiently illustrated by reference to the monograph by Dodge (1927), in which, purporting to deal with variability of response in human beings, there are descriptions of the fact that the same individual is not always alike in response to "the same stimulus." Dodge urges study of "progressive modifications of response" —but these modifications (of reaction-time) are not treated as functions even of psychologically or biologically significant "time"; and the assumption that the "stimulus" is the same because the same change in the intensity factor of some form of energy external to the organism has been employed as a means of excitation is palpably exposed to the possibility of serious error. The fact that the *variation* in response, at any one period of time, is described statistically, does not enable us to discuss parameters of the variation as a function either of the conditions of excitation or of any change in the organism. Yet it is this type of procedure, namely, the mere recording of differences (qualitative merely, or between averages), with or without characterization of variation about average values of "response," which has usually been employed in the investigation of variability. The methods of correlation do not supply what is needed, since they do not provide means of expressing *variation* as a function of some independent variable, but merely state in terms of pure number the degree of association between mean values of co-varying attributes.

To describe the *variability* of a response, then, it should be necessary to characterize not the mean value of the response as a function of something else, but the variation as such a function. In this way it may be possible to gain some insight into the nature of variation in particular cases. The variability of response might then be found to be in these cases far from lawless. The importance of such findings would be considerable. Upholders of the view that behavior is essentially unpredictable, that is, not directly a function of knowable conditions, inevitably trade upon the margin of variation supposed to be typical of organisms as contrasted with machines. If variability can be shown not to be lawless, the supposed logical foundation for this type of obscurantism disappears.

Moment-to-moment changes in the behavior of animals have practical consequence in at least two ways, from the standpoint just described. (1) They frequently interfere with or complicate attempts to analyze the mechanics of irritability and response by way of quantitative tests of relations between excitation and reaction. As a source of such trouble, intrinsic variability has usually been regarded as an inescapable nuisance, perhaps justifying superficial control of external variables; but since "errors of observation" may influence end results in a multiplicative way, and *not* by a principle of exclusive limitation, this view is, practically, unjustified. (2) On the other hand, it may be possible to utilize this very variability, provided it can be expressed decently in quantitative terms. In the illustration of the basic notion employed we must for the present be content to rely upon relatively crude statistical methods, because the existing data do not justify the labor inseparable from the execution of certain obvious refinements. This in no way restricts the generality of the conception. It is unfortunately necessary to point out that the data of psychophysics as yet supply very scanty material for the purpose—although it is in this field that they might be most expected to occur. However this may be regarded as a criticism of the usual psychophysical methods, it is in no sense an adverse criticism of the procedure proposed.

The practicable characterization of the variability apparent in the responses of different individuals of the same general sort, under external conditions physically similar throughout, may be based upon one of several considerations, applicable in different instances. We might deal with the standard deviations of the measurements, or with the relative standard deviations of the measurements (as percentages of the means), or, again, with the absolute ranges of scatter of the data. Presumably every case would have to be examined on its merits, to permit decision as to whether "errors of observation" predominate in the variability of the data, or variation of an intrinsic sort. In general, it can be foreseen that it is desirable to obtain measurements upon the *same individuals* at each of a number of values of the experimentally controllable variable, thus keeping the foci of origin of variation the same as to number and as to kind. And it is also clear that the numbers of measurements made should likewise be uniform, else the computed statistics will not have equal weights. If in any series of measurements such as we might desire to use there are present two different sorts of "variation," respectively, those for which uncontrollable "errors" are directly responsible and, on the other hand, those for which the inner variability of the organism is the cause, the disentangling of the two will in general depend upon keeping the numbers of observations constant at each value of the chosen independent variable.

Without treating in detail the methods whereby such observations have been obtained, we take as illustration data from experiments already cited on the geotropic orientation of young rats. Each series of measurements involved the recording of a certain fixed number of orientation angles (no rejections) exhibited by each individual in a given homogeneous lot of young rats, under standardized conditions, at each of a number of inclinations of a surface on which creeping occurred. The mean values of the angles of the upwardly directed paths of steady creeping (θ) were found to be, as an empirical approximation, linearly proportional to log sin α, where α is the slope of the surface.

For the young rats (and for young mice) it was found that the variability of the measured orientation decreased as the extent of upward turning became greater. The interpretation proposed turned upon the assumptions: (1) that upward orientation involved an adjustment of the axis of the body such that the pull of the mass of the body on the legs induced tension excitations during progression which at the definitely oriented position became equivalent (i.e., equal, within a threshold difference) on the two sides of the body; and (2) that increasing the effective pull of gravity, proportional to sin α, involved greater total excitation (measured therefore by the angle θ), and at the same time proportionately interfered with the disorienting or otherwise competing effects of other sources of excitation. If in any one series of measurements the variation of response is due to a set of conditions, statistically uniform at different values of sin α, which may exert minor influences upon the angles of orientation, and if the several individuals used are strictly comparable, and if the foregoing assumptions (1) and (2) are valid (particularly 2), then the variation of response (θ) should be proportional, at each value of sin α, to the ratio V/θ, where V is the integrated influence of the minor orienting "errors" (including errors of recording and measuring the orientations). Hence the various assumptions can be tested by ascertaining if σ_θ/θ

is influenced in the same way by increasing sin α as in the extent of orientation (θ) itself (70, 75). In the papers previously cited it is shown that this is the case, since σ_θ/θ declines rectilinearly as log sin α increases. Since in any one series of observations the number of individuals used (N) is constant throughout, and the number of observations made on each (n) is the same at all values of α, the fact that the distribution of θ is skewed (76) at each magnitude of α is in no sense a disadvantage. If it were necessary to correct for skewness, this could be done (a) by recognizing the logarithmic nature of the skewness, or (b) by dealing with the symmetrical frequency distributions of cos θ, which is a linear function of sin α. The standard deviation *of the mean* (σ_M) is used, not the mean standard deviation of the single observations (σ_1), since we desire to compare *mean* θ's.

As a qualitative proof of the legitimacy of the view thus far developed it may be pointed out that: (1) as θ increases in these series, σ_θ decreases absolutely as well as relatively; (2) when inhomogeneous groups of individuals are purposely used, the formulation fails in a predictable way (76); (3) when different individuals are used at different values of the slope θ, the relationship of σ_θ/θ to log sin α is no longer so exact, though it remains roughly proportional (140, 99). Quantitative proof is obtained in another way.

In Figure 25 there is representation of a series of measurements of this kind. In this case reliable measurements of θ were obtainable at slopes (α) between $\alpha = 20°$ and $\alpha = 70°$. On the ordinate axis is given $P.E._\theta$ as percentage of θ, i.e.,

$$0.6745 \cdot \sqrt{\frac{\Sigma v^2}{n\ (n-1)}}$$

where Σv^2 is the sum of the squared deviations of the individual observations from the mean of all taken together; n is the number of readings made at each slope (in this case, 40); the use of $P.E._\theta$ rather than of σ_θ merely changes the size of the ordinary units, and does not alter the slope of the line suitably computed (data in 76). It may be mentioned that the probable error of the plotted quantity (i.e., of the "coefficient of variation") also declines, and in the same way.

In such a graph as that in Figure 25 the area under the fitted line, between the working limits of the observations, measures the total "variation" of the observations. It is clearly composed of two parts. The lower portion is that uninfluenced by the increase of sin α, and includes errors of observations and of recording. With trained observers, and with observations taken within a limited interval (but in such a way as to avoid fatigue effects), it should be constant in any one series of the kind earlier specified. The upper section of the area in question is that proportion of the total variation which is subject to control according to the magnitudes of the independent variable concerned in the experiments, namely, by log sin α. This will clearly involve, also, dependence on the number of different individuals used, and on the (constant) number of readings taken with each. We are interested in the connection between the *variation of the response* measured (i. e., by σ_θ at any value of α) and the magnitudes of the exciting stimulus, in this case sin α. Having chosen a formulation which gives a sufficiently rectilinear plot, we measure this dependence of σ_θ on sin α by means of the slope: $\Delta\sigma_\theta/\Delta \log \sin\alpha$. This should give an index of *variability* free from the effects of "errors of measurement," since it concerns the way in which the controllable variation in response is modified by the variable which determines the response. We have elsewhere pointed out (76) that the comparison of two groups of individuals as regards *amounts of response* cannot, in general, be made merely by comparing their responses at *one* value of the intensity of excitation. For analogous reasons, in the present connection it is to be stressed that relative variability of response cannot be examined by contrasting *variation* of response in two individuals or groups, with a single value of the controlling variable, unless it is demonstrated that the variation as measured contains no element which is not modified by the variable in question. It should be clear that since $\Delta\theta/\Delta \log \sin\alpha$ is practically constant, in these series of measurements, plotting σ_θ/θ against θ will also give straight lines; but for the present we desire to keep the dimensions of our unit of variability ($\Delta\sigma_\theta/\Delta \log \sin\alpha$ with N and n constant) clearly in view, and hence we shall in the further examples use chiefly plots of σ_θ/θ against log sin α.

Such formulations might be examined in terms of the theory of probability and of statistical constants; or they may be approached in an experimental way with the intention of discovering whether the constants they contain have a "real," that is (in this case), a biologically real significance.

The measure of variability of response is one which uses such units as express the change in *variation* of response per unit change in the controlling independent variable. Under the conditions we have described, it will clearly depend upon the number of observations at each value of the controlling variable. It will also depend upon the number of "centers or foci of variation," that is, on the number of individuals used. If the individuals used are not strictly comparable one with another, the *controllable* variation in response must proportionately decrease. In the example to be used, uniformity of material has been assured by the selection of genetically stabilized groups or races of rats. The points here stressed can be illustrated and shown to be essential from the results of tests with inhomogeneous groups of individuals. (Since what we desire to measure is the variability of "inner" origin, the ideal procedure would be to work with single individuals; for technical reasons that is impossible in the present case, as the time within which observations remain comparable is limited while "fatigue" and other complications must be avoided.)

With a race termed *A*, three lots of individuals were employed. The first two sets, *A*I (1-6) and *A*I (7-12), were two litters from the 10th inbreed (brother × sister) generation of a line previously inbred "loosely"; Series *A*II (1-5) was from a litter of five individuals in the 12th generation of closest inbreeding. In Series I, 8.3 observations (on the average) were made with each rat at each value of the slope a; in Series II, 15 observations were recorded on each. In terms of the preceding analysis, if two series of comparable observations are made, one with N individuals and 20 observations on each, the other with $2N$ individuals and 10 observations on each, the measure of the intrinsic variability should in the former case be one-half as great; that is, the slope $\Delta \sigma_\theta/\theta / \Delta \log \sin \alpha$ should be twice as great as the first case. The measure of variation should be proportional to $\sqrt{1/n}$, where n is the number of observations at such value of a; it should also be proportional to $\sqrt{N}$, where N is the number of very similar individuals serving as separate foci of variation of response. In the example cited, then, the variability should be

$$a\sqrt{\frac{N}{n}} \text{ and } \sqrt{\frac{2N}{10} \mathrm{x} \frac{20}{N}} = \sqrt{4} = 2,$$

or the variability should be *twice* as great with twice as many individuals, if the number of readings made with each individual is half as large. An experimental test of this is given in Figure 25.

Within certain limitations, this method can be generalized. Series *A*I was run in two stages, using two litters of 6 individuals each. The slopes $\Delta \sigma_\theta/\theta / \Delta \log \sin \alpha$ should be the same for each group. The only known difference between litters 1-6 and 7-12 was that 7-12 was a day older and a little heavier. The slopes are not exactly equal, but in the ratio 0.83. If we compare these data with those from Series *A*II, we expect the slopes to be, for Series AI (1-12) and Series AII (1-5), in the ratio

$$\sqrt{\frac{75}{100}} \times \sqrt{\frac{5}{12}} = 1.34.$$

The ratio found is 1.46. Between Series AI (1-6) and Series AII (1-5) we expect, on the same basis, a slope ratio of 1.29; that found is 1.28. Agreement of this kind is, of course, an excellent test of the homogeneity of the material used, of the uniformity of the technique, and of the unbiased nature of the observations.

This mode of calculation can be extended to the comparison of different races as regards variability. If two races of young rats are essentially similar with respect to properties which may influence the expression of θ as a function of α, under the conditons of these experiments, then we expect the slopes of the lines connecting σ_θ/θ with log sin α to be in the ratio already given,

$$\sqrt{\frac{(N_1)\ (n_2)}{(N_2)\ (n_1)}}$$

where N_1, N_2 are the numbers of individuals of races 1, 2 concerned in the tests, n_1, n_2 the respective numbers of observations at each magnitude of α with each individual (constant at every α for each race). Deviations from the expected ratio would provide a measure of unlikeness, as regards variability of the response, in the two races. In other words, we would compare the values of a "variability number"

$$\left(\frac{\Delta\ \sigma_\theta/\theta}{\Delta\ \log \sin \alpha}\right)\left(\sqrt{\frac{n}{N}}\right)$$

having the dimensions of "the change in the mean root-mean-square deviation from the mean, per individual, as a percentage of the mean, per unit change in the controlling variable." Within certain limits this *variability* number should be independent of the number of observations and of the number of individuals used, provided these are the same in any one series. For race K and race A, "duplicate" experiments show agreement as close as can be expected. Differences between the several races are clear. It is to be remembered that the slopes $\Delta\ \sigma_\theta/\theta / \Delta \log \sin \alpha$ as obtained graphically are subject to errors of curve-fitting which may make appreciable differences in the slopes, and thus in the final figures.

If an additional load is carried by the legs during progression, as when a small brass weight of suitable form is properly fastened to the back of the young rat, the curve connecting θ with α is distorted in a particular way; its low-α segment is pushed upward (higher values of θ). If the *variability* of the geotropic response is inversely determined by the intensity of excitation, then our "variability number" ought to be reduced; the values of the function $\Delta \sigma_\theta/\theta \Big/ \Delta \log \sin\alpha$ should be less, with samples of the same size. The comparison is fairly made, however, only when the variations are compared at equivalent values of the total response, and we have already pointed out that the angle θ is a measure of the total excitation. The increase of θ at the lowest working value of the slope α, however, measures the increased efficiency of excitation due to the presence of the added weight. In the present case the increase, at $\alpha = 20°$, was from $\theta = 34°$ to $\theta = 51°$. With n and N the same, then, we expect the slopes $\Delta \sigma_\theta/\theta \Big/ \Delta \theta$ to be inversely in the ratio 34/51 or 1 : 1.56; actually they are found to be in the ratio 1.6.

A case in which an organic discontinuity in performance is checked by of behavior of the variability function is illustrated in Figure 33.

Certain technical difficulties may attend the investigation of variability in the case of responses or reactions simultaneously controlled by more than one form of excitation, but hints already available show that the attempt is profitable not alone as a check upon the validity of measurements but also as supplying data for the ultimate evaluation of the mechanics of variation and "choice."

We have not sought to imply that in all cases variation of a specific form of behavior can be expected to change in the same way as the quantitative reaction or response itself. It is true, however, that data from certain additional sources—from the study of judgment of lifted weights, for example—can be analyzed in these terms. But among a number of further instances it is worth while to choose several in which the variation of "reaction" of performance is independent of the condition of the performance itself. In any particular case it may remain a problem as to whether the standard deviation or the *relative* variation (σ/mean) is the proper measure of variation, and this must be decided by additional evidence and in part by the nature of the inquiry it is desired to make. Lashley (1915) found that with increasing skill developed in successive equal practice periods the range of scatter of "shots" in archery decreases, but that the relative variation remained practically constant (the same number of shots being made in each period). In such cases it is tempting to conclude that the mechanism of variability is one which is independent of that governing the change in accuracy due to experience. Similar conditions appear in certain other data relating to learning, memorizing, and forgetting. In the early excitements connected with the popularization of biometry a similar problem arose as to the proper index of variations in the measurements of organisms (cf. Duncker, 1900), namely, the question as to whether (1) the standard deviation of a character, or (2) the ratio between this statistic and the mean (or median) of the character is a proper

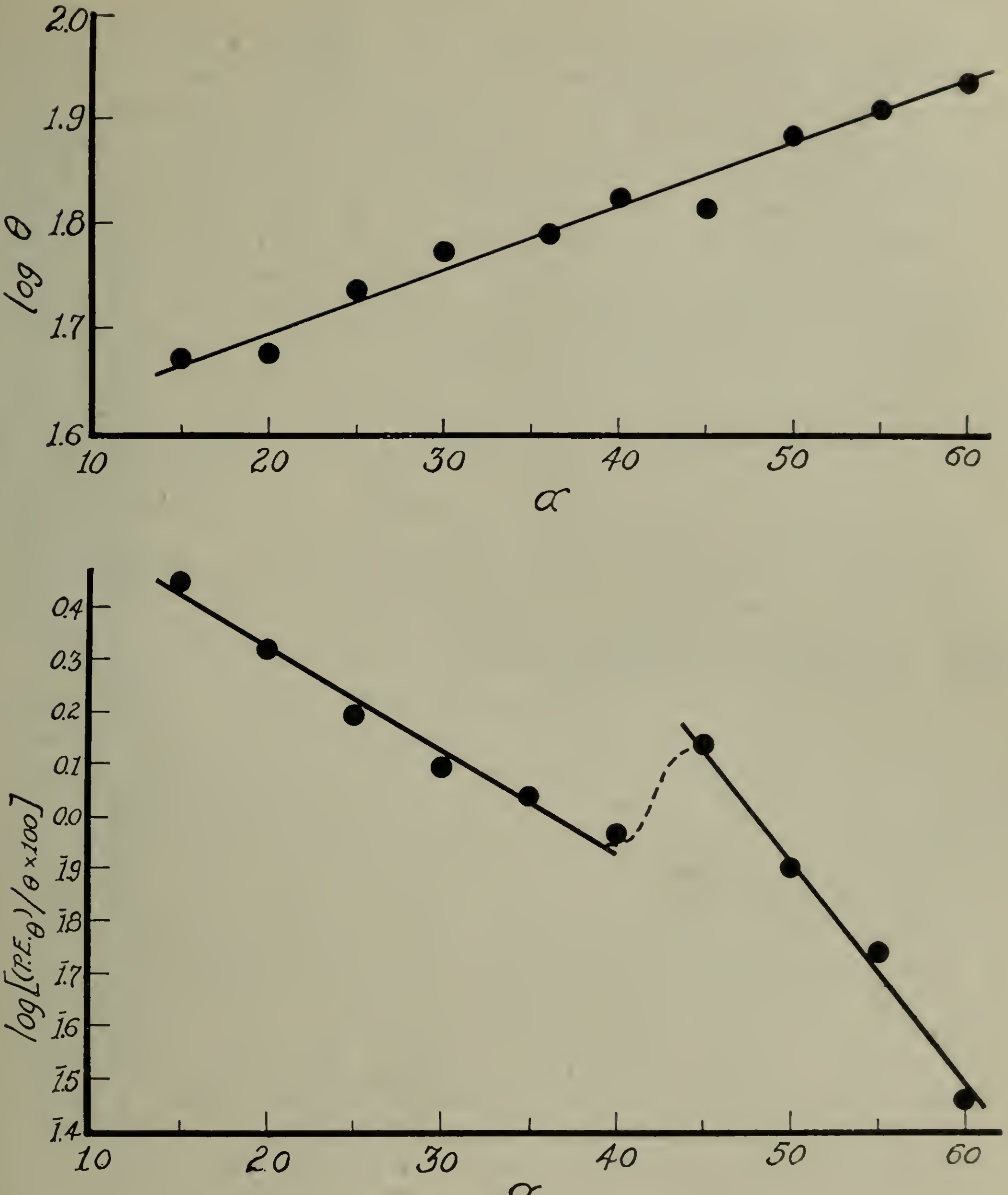


FIGURE 33

Data from One of Several Series of Experiments by Dr. Morgan Upton on Geotropic Orientation of Young Guinea Pigs

In this case the method of support upon the inclined surface (α = slope) differs from that in rats, the body being free above the surface. Empirically, log θ *vs.* α gives in this case a nearly straight line. The cusp evident at $\alpha = 45°$ is real, however, and is correlated with an easily observable change in the method of progression at slopes above this magnitude. Its reality is made evident when the variability function is considered. In the lower graph the logarithm of the relative *variation* of θ (n here = 80) is plotted against α; the unmistakable discontinuity is apparent at $\alpha = 45°$, marking the transition to a new set of conditions at greater slopes which affect both θ and its variations as functions of the slope.

index of its absolute variability. Verschaeffelt (1894) assumed the second of these alternatives, and Davenport (cf. Brewster, 1897), Pearson (1897), and others also assumed the relation between σ and the average for a character to be similar to that between the errors of measurement by successive applications of a foot-rule and the length measured. Duncker (1900) believed that there is no relation whatever between the average magnitude and the variability of a character, since in his view the causes of variation are to be considered essentially different from the factors determining total size. The argument included the notion that homologous characters in allied forms ought to show similar indices of variability, but not necessarily equal average values. Certain series of measurements (as of rostral spines in *Palaemonetes*) show variation in a way supporting this view. In general, a *coefficient of variation* of the order 5 to 10 *per cent* is found for continuously varying biological characters; for seriations of integral variates C.V. tends to be larger, and more irregular, while σ is relatively more constant. The problem assumes a slightly different form, however, when the intention is to investigate variation as a function of some controllable quantity external to the organism. We can illustrate this briefly by means of data drawn from studies on the relations between temperature and the speeds or frequencies of certain processes in organisms.

We choose data of this sort because it again shows a way in which variation of performance, instead of remaining a handicap, can be made to prove something in a useful way. To do so requires experimentation with the utmost care for the control of all recognizable external sources of difference in the mean values of the measured quantities, other than provided by the single variable one uses for the investigation. The control of this one variable, temperature, must be as exact as possible. Where the temperature of the organism differs from that of the surrounding medium it must be measured by special means. Into the technical side of such operations we need not go. If the frequency of the heart-beats be measured, let us say, in caterpillars of the same stock, at each of a good number of temperatures, over a range of temperature within which change of temperature produces no irreversible effects, there is found a definite and smooth relationship between temperature and frequency. The observations are secured in such a way that the time is always measured for 10 beats of the pulsating vessel. A well-known law, with adequate theoretical underpinning, relates the velocities of chemical reactions proceeding at measurable speeds to temperature. It states that the velocity constant is proportional to the exponential of $-\mu/RT$, where R is the gas constant, T is the temperature on the Kelvin scale, and μ is a quantity having the dimensions of "energy of activation" of the molecules whose activation governs the rate of the process. It happens that the same equation applies to the biological situation. The problem then arises, does the quantity μ have, in such instances, a *specific* significance? It is possible to suggest a good many reasons why it might not have any specific significance. Protoplasm is colloidal; temperature in-

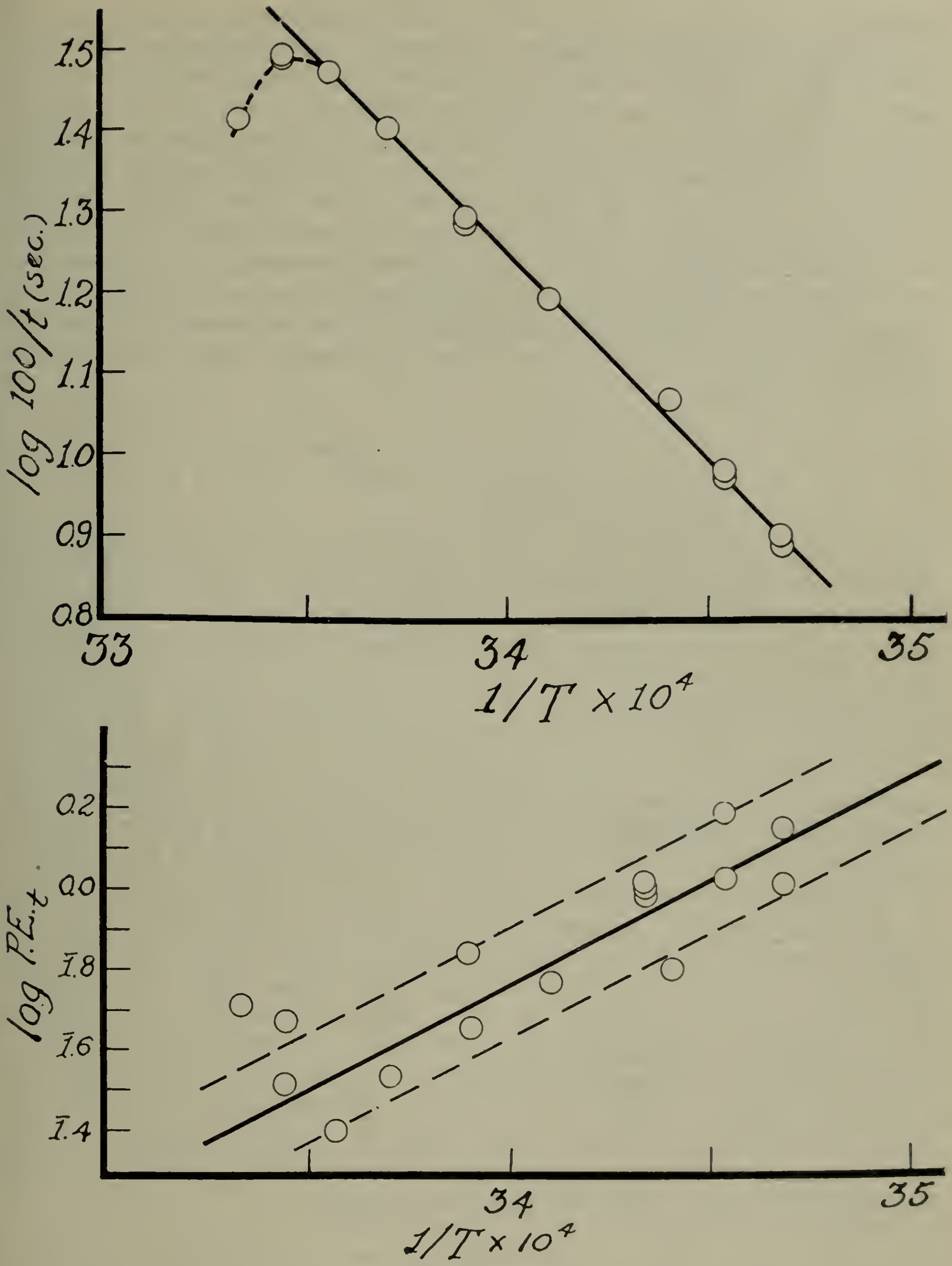


FIGURE 34

To Illustrate Certain Features of *Variability* of Performance in a Living System

A typical example of the relation between the frequency of breathing movements and (internal) temperature in a young mouse. The points on the upper curve are averages of 40 readings each. Log *frequency* declines rectilinearly with increase of $1/T$, where T is the temperature Kelvin. The log of the probable error of the mean *time* for 10 *movements* increases in the same manner. Hence the relative variation of the mean frequency (or time for a constant number of movements) is independent of the temperature. A temperature of 25° C. is "critical" for the mice of this particular strain. This is shown both by a decline in frequency of movements, and by the concurrent increase in variation. (Data from experiments by Dr. G. Pincus.)

fluences its viscosity, and the like. Such suggestions are more in the nature of apologies for avoidance of rigorous experimentation, however, than anything else, and appear to be pretty certainly based upon misconceptions of the rôle of viscosity and diffusion in microscopic systems, because it would then still remain mysterious why relations of the type illustrated in Figure 34 are in fact the rule (with indeed but few exceptions). We are not now concerned with this problem of the possible or even probable specific significance of magnitudes of the constant μ in biological systems, which the evidence, however, clearly supports, but with the method of using the variability of the measurements. The plot in Figure 34 shows that the "time for ten movements" is not absolutely constant. Neither does it vary at random. On a logarithmic plot of the function

$$\frac{1}{\text{time for 10 movements}} = Ae^{\frac{-\mu}{RT}} + \text{const.},$$

that is, in terms of the physicochemical theory previously mentioned, the observations tend to lie strictly in a band with parallel edges. The latitude of variation is then such that the highest frequencies of movement are a constant times greater than the lowest frequencies (77). This means that the latitude of variation is a constant percentage of the mean frequency, *and is independent of the temperature.* Since the latitude of variation in such experiments is far outside the limits of any possible "errors of measurement," and certainly represents *organic* variation, this means that the physical condition of the milieu in which the frequency of movement is determined is not significantly altered as a function of temperature. The *variability* of the frequency, as a function of temperature, is therefore zero, or indistinguishably close thereto (52, etc.). One is thus allowed to conclude that change of temperature (within specified limits) does not alter in any relevant way the physical properties of the system in which there takes place the process determining the expression of the mean frequency of movement as a function of temperature. It would be difficult to picture a system physically changed by change of temperature in a manner influencing the mean frequency of movement by alterations in protoplasmic fluidity, for example, which would allow the relative variability of the frequency to stay constant. The instances cited do not stand alone; they are typical of a fairly large number now available.

VII

Granted that under certain conditions it is possible to account quantitatively for elementary manifestations of geotropic and of phototropic orientation, it is desirable to exhibit in a summary way specific applications of the analyses. These consist, in a sense, in testing one's ability to predict quantitative features of response in situations such that both phototropic and geotropic orientations are in evidence. Conceivably this sort of inquiry might ultimately be pushed to include the nomographic representation of determined conduct when such variables as gravity, light intensity and wave-length, galvanic current, temperature, oxygen tension, time, and other conditions can be controlled. Thus far, the data do not exist for such an extension, and the simpler instances about to be mentioned are, from an experimental standpoint, difficult enough.

With young rats it was possible to expect (73), from previous acquaintance with the geotropic orientation and phototropic orientation (in the case of animals with unopened eyes), that the intensity of light required to effect a definite phototropic counterbalancing of the geotropic orienting influence should be a power function of the sine of the angle of inclination of the creeping plane (Figure 35). In addition, it was possible on the same basis to account for the changing variability of the measured values of I as α was increased (Figure 36). To this case may be added that in which the compounding of galvanotropic and photic excitations was considered. It is of interest to deal at greater length with a case where time enters as an additional variable. This will illustrate not only the general problem of orientation in compound fields but also the way in which the results of this procedure lend themselves to a more penetrating treatment of mechanisms of excitation.

It is necessary to stress the fact that organisms and individuals exhibiting the response desired must be selected if properties of the response are to be investigated. The experiments involve the effect of phototropic excitation upon the direction of gravitationally oriented creeping of the slug *Agriolimax*. The operative problem first consists in securing a measure of the orienting effect of light, with the gravitation influence held constant. In darkness a negatively geotropic *Agriolimax* upon a vertical surface creeps straight upward. When light (horizontal rays) falls upon one side of the animal, previously dark-adapted, it continues to creep upward but at the same time away from the light, the path departing from the vertical by an angle β (Figure 37). Time enters as a variable, since photic adaptation is fast. While it would be preferable in some respects to measure at each chosen moment the intensity of light necessary to produce a constant deviation of the path from the vertical, it has been easier to measure the deviation itself as a function of time (141). Two forces are apparent in such experiments. The light leads to photonegative orientation, and the geotropic stimulus forces the slug to creep upward. If, during exposure, the test light be removed, orientation is at once upward. If, on the other hand, the inclination of the creeping plane be decreased, orientation is more completely away from the light. Therefore the path actually followed depends upon the effects of concurrent continuous stimulations of the two kinds. The geotropic stimulus may be regarded as constant. Then the angle β depends on the relation $\tan \beta = \frac{L}{g}$, where L is the phototropic vector. We may

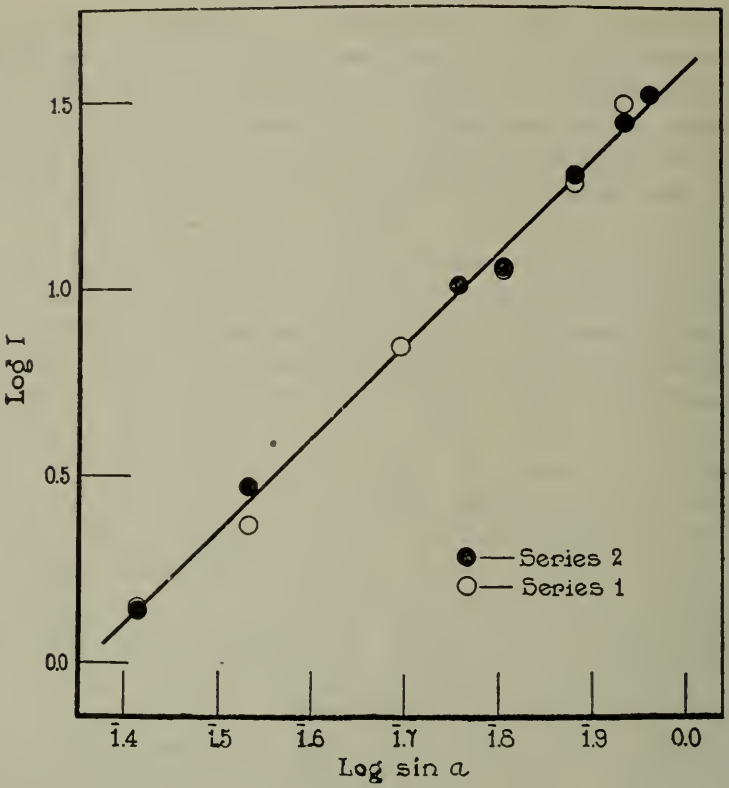


FIGURE 35

The intensity of light required to inhibit geotropic orientation of young rats varies with the inclination of the creeping surface, in such a way that Δ log I/Δ log sin α is constant. For Series 1 each point is the mean of 15 determinations; in Series 2, of 20 determinations. (From *J. Gen. Physiol.*)

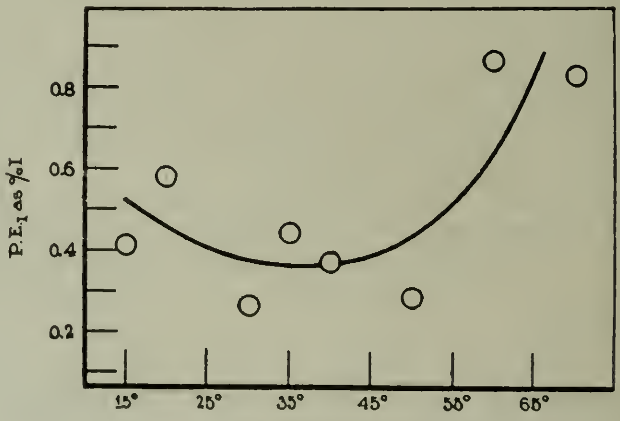


FIGURE 36

The variability of the intensity required to counterbalance geotropic excitation passes through a minimum as the angle of inclination of the creeping surface is increased (see text). (From *J. Gen. Physiol.*)

assume that at any moment the photic irritability is proportional to an amount of photosensitive material S, only very slowly produced by the processes underlying dark-adaptation, and that the photolysis of S adheres to the kinetics of a first-order reaction (88-90). This means that over any short interval the decomposition products formed from S, which are responsible for excitation, will be proportional to S_t and thus to S_0e^{-Kt}, when S_0 is the amount of S at complete dark-adaptation. Inserting this relation in the previous equation, tan β should be proportional to e^{-Kt}, and log tan β should be directly proportional to $-t$. Moreover, the rate coefficient of light-adaptation, K in the preceding expression, would be expected to be practically proportional to the logarithm of the intensity of the light used to produce adaptation. Both of these expectations are verified (Figures 38, 39). They are further checked by the orientation effect with forms only *slowly* adapting to light (142) as in *Limax,*

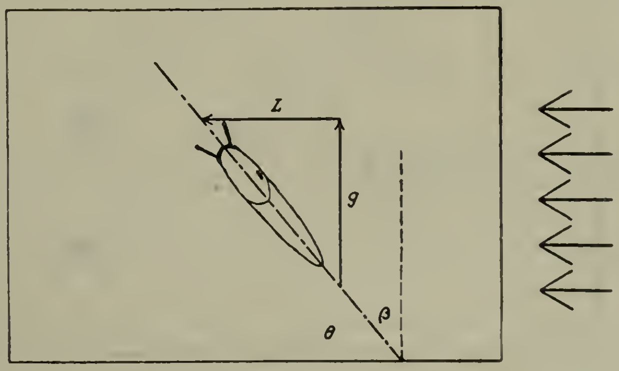


FIGURE 37*a*

Orientation of negatively geotropic and negatively phototropic *Agriolimax* on a vertical plate with light from the right, —to indicate terminology of the phototropic (L) and geotropic (g) vectors, the angle of orientation (Θ), and its complement (β). (From *J. Gen. Physiol.*)

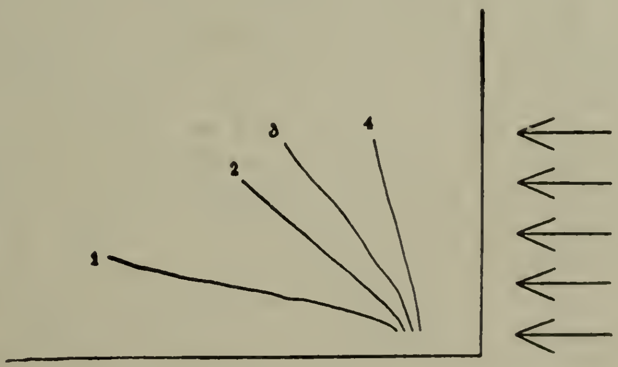


FIGURE 37*b*

Successive Trials Made by Initially Dark-Adapted *Agriolimax* on a Vertical Plate, Light from the Right, at Intervals of 1 Minute (Mid-Points)

(This gives the form of the records as available for analysis; actually the trails all began at the same point.) (From *J. Gen. Physiol.*)

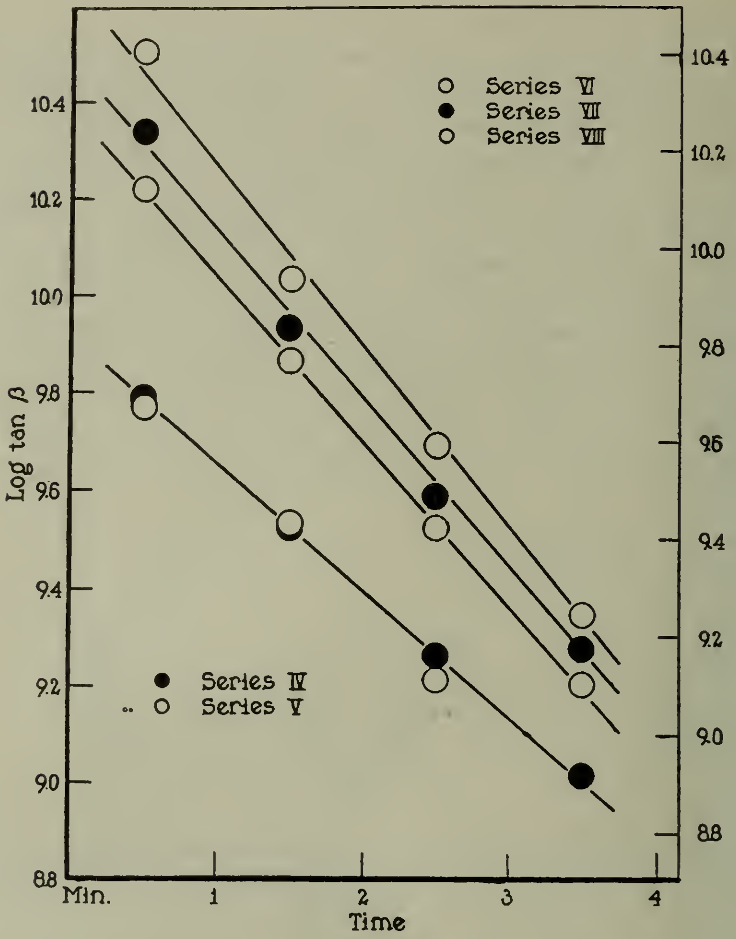


FIGURE 38

The course of light-adaptation in five series of experiments, at two different intensities: for Series IV, V, 16.32 f.c.; for Series VI to VIII, 73.69 f.c. (From *J. Gen. Physiol.*)

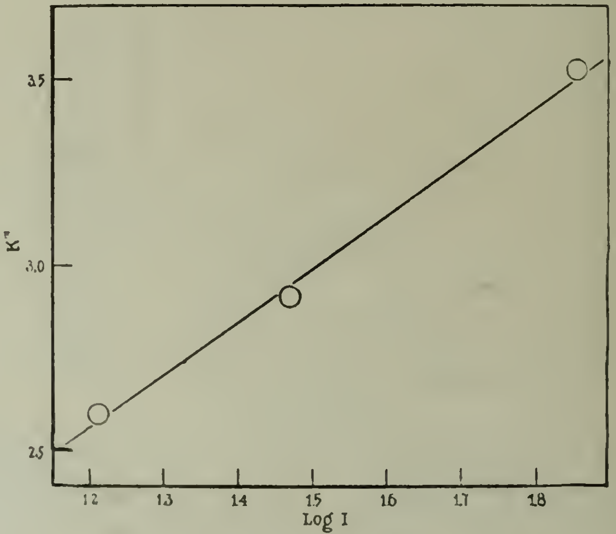


FIGURE 39

The rate of light-adaptation of *Agriolimax*, as measured through the opposition of photic and constant geotropic orienting forces, increases directly as the logarithm of the light intensity. (From *J. Gen. Physiol.*)

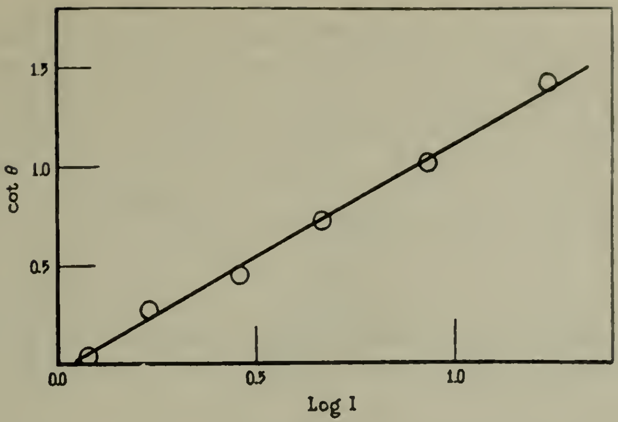


FIGURE 40

The cotangent of the angle of inclination of the body of *Dixippus,* on a vertical surface with illumination from one side, is a linear function of the logarithm of light intensity. (From *J. Gen. Physiol.*)

(Figure 40), even where the vectors g and L are not at right angles (Figure 41).

Since β decreases with time, it might be suspected that the slug "learns" to overcome the effect of the light. This would suppose that the concomitance of geotropic and phototropic excitations produces, centrally, a "block" for photic impulses (or else, enhances the geotropic effect). This would be interesting in itself, if verifiable, since the process clearly obeys a definite law. The alternative interpretation is that the photic adaptation is direct, and occurs in the eye. By means of experiments in which the animal creeps on the vertical plane only during the brief interval required to measure β, the exposure to light being continuous, it was shown (79) that the measurable adaptation does not depend upon the simultaneous competitive action of the geotropic stimulation (Figure 42); it must therefore be regarded as located in the eye.

The notion that excitation by light is a photochemical matter, and that the maintenance of photic reactivity and its recovery after exhaustion may be in some fashion connected with a reversible photochemical receptor system, seems to have been held in a very general form since Hering. Growing knowledge of the chemical effects of light made it attractive to translate such ideas into photochemical terms. An early attempt to treat the question precisely and specifically, but with some important deficiencies, was made by Müller (1896). To explain the simultaneous exhibition of capacity to react to a sharp decrease in light intensity and also to be excited to phototropic movements by the continuous action of light, as seen in the integument of certain holothurians, it was suggested that the same photochemical system, if part of a reversible reaction of a certain kind, might be conceived to serve for both modes of stimulation (42). Hecht has been able to achieve the step-by-step proof, in a very satisfying and convincing manner, that a particular type of reversible photochemical system underlies photic excitation of *Mya, Ciona, Pholas,* and the human

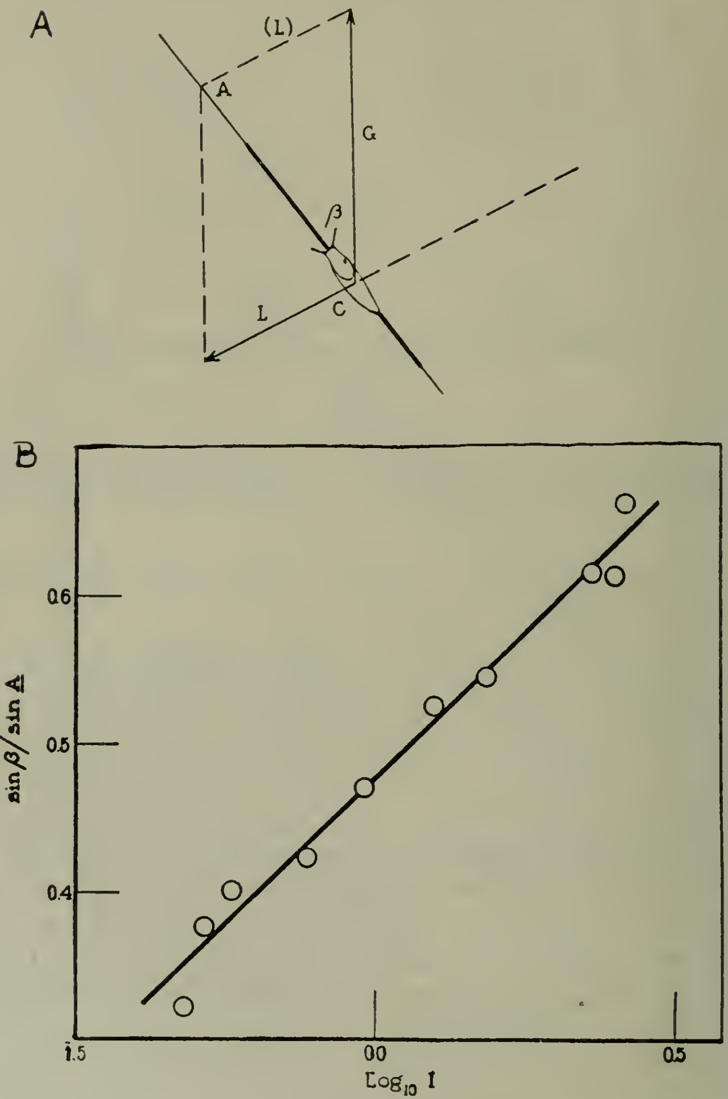


FIGURE 41

A. A slug (*Limax*), previously dark-adapted, is creeping on a vertical plate of ground glass which has been moistened. It is illuminated at an angle from one side by light from a small condensed-filament lamp. To avoid effects of the observer's breathing currents, the path of the slug is observed from the other side of the plate. At the point *C* the slug is under the influence of an orienting field containing two significant vectors: a gravitational vector *G*, which is assumed constant; and a photic vector *L*, of magnitude assumed proportional to log *I*. *I* is *measured* at *C*, with a photometer giving illumination in foot candles. The direction of progression is given by the heavy line. Then if the assumptions are valid

$$L = K \log I = \left(\frac{G}{\sin A} \right) \times (\sin \beta);$$

or

$$\frac{}{\sin A} = \log I.$$
$$K' \sin \beta$$

The plot in Figure B shows that the equation is satisfactorily obeyed. The residual variation is about the same at all points (*ca.* ± 4.0°).

eye. The properties of this system permit the prediction and the rational quantitative interpretation of new and unexpected characteristics of the stimulation process. A major property of the underlying mechanism is adequately expressed (85, 88) in the following paradigm, where S represents primary photosensitive substance, P and A two of the products of its photolysis: $S \underset{\text{dark}}{\overset{\text{light}}{\rightleftarrows}} P + A$. The system is reversible, and with time, under continuous constant illumination, there is established a stationary state ("equilibrium"). The "dark" process, of which the kinetics serve to describe the course of dark-adaptation, is definitely bimolecular

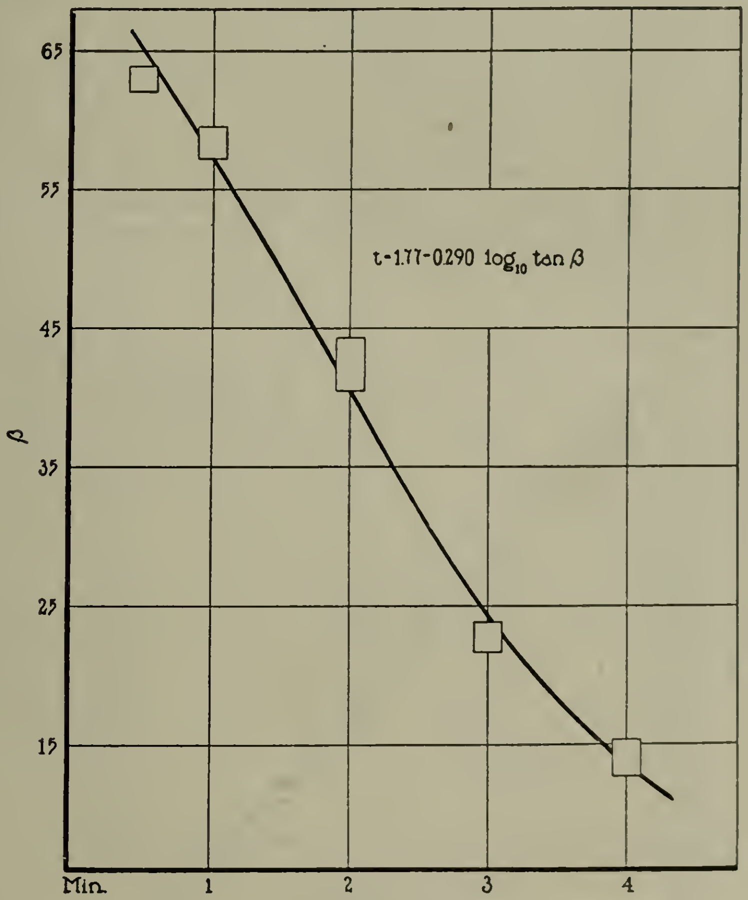


FIGURE 42a

The angle of orientation of *Agriolimax,* the departure (β) from a vertical path upon a vertical plane as forced by the phototropic effect of light from one side, after increasing periods of exposure to illumination of 29.4 f.c. The slugs were creeping upon a *horizontal* surface except during the half minute required to obtain the orientation trail. The observations are plotted as bars centered on the means, with height = 2 P.E. The curve is that of the equation

$$\text{time} = 1.77 - 0.290 \log_{10} \tan \beta,$$

and the agreement with the observations shows that, as previously found (141), log tan β decreases linearly with time. (From *J. Gen. Physiol.*)

(91), with high temperature characteristic; the "light" reaction is of first order, and negligibly influenced by temperature (86).

These relationships have been established most elaborately by means of experiments with organisms in which the speed of response following the delivery of a measurable quantity of light gives data necessary for the quantitative treatment, and in which light-adaptation is rapid and photic excitation therefore necessarily discontinuous. To carry over such considerations to the analysis of continuous excitation, as in phototropic stimulation, it has been necessary to find phototropic organisms in which the rate of light-adaptation is measurable, neither inconveniently rapid nor too slow. The photic adaptation of *Agriolimax* has been followed by causing the phototropic influence of the light to work against the "brake" provided by a vectorially constant excitation of a different kind, namely, that due to gravity, acting at right angles to the phototropic vector. The function of such a "brake" in contributing to the significance of the measured orientations has been discussed previously in relation to circus movements (62). Data necessary for treatment of the light-adaptation are obtained from measurements of the resultant angles of orientation as related to time and to intensity (141). Such measurements are possible because the rate of light-adaptation, especially at temperatures in the

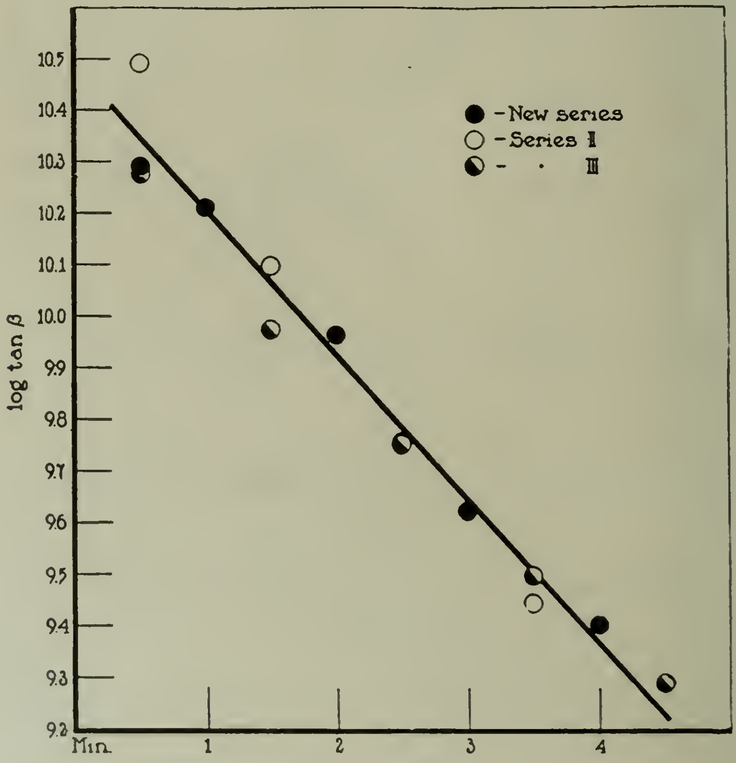


FIGURE 42*b*

Data from two earlier series (II, III) of experiments are plotted together with the present series, to demonstrate the agreement in the extent and in the rate of photic adaptation, and thus their independence of continuous concurrent geotropic excitation, as practiced in series II and III. (From *J. Gen. Physiol.*)

neighborhood of 15° or slightly above, is so very much faster than that of dark-adaptation, as adequate tests showed. It was easily established that the rate of light-adaptation is very little influenced, if at all perceptibly, by temperature; and that its course follows that of a first-order reaction in which the "velocity constant" is a linear function of the logarithms of the light intensity, as already indicated. These phenomena therefore

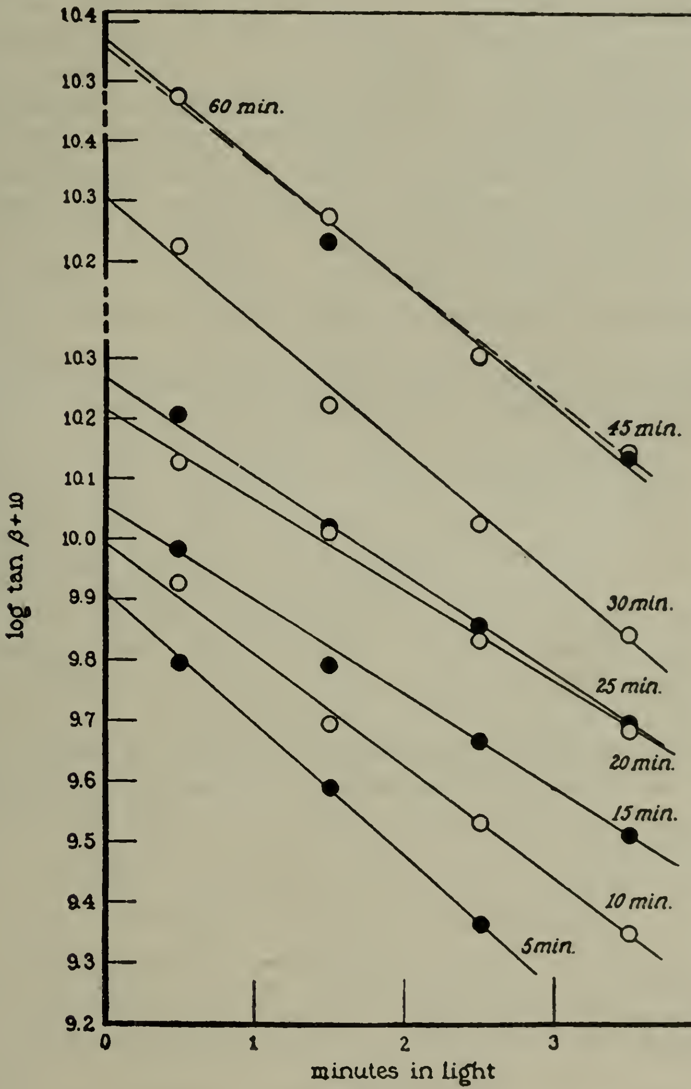


FIGURE 43

The angle (β) of orientation of *Agriolimax* away from the vertical path on a perpendicular plate, as enforced by exposure to horizontal light of 29.4 f.c. on the right side, at successive minutes during exposure, following increasing residence in darkness (5 to 60 minutes) of slugs initially light-adapted. Time of exposure plotted against log tan β gives graphs sensibly rectilinear. The intercepts of these graphs on the ordinate axis give a means of estimating the photic excitability at the very first moment of exposure to light (see text). (The ordinate axis has been shifted vertically, for "30 min." and for "45 min." and "60 min.," to obviate overlapping.) (From *J. Gen. Physiol.*)

parallel, in an exact manner, the properties of the "light" reaction already discussed.

The interpretation of dark-adaptation requires measurements of the photic excitability at successive intervals following exclusion of light from individuals previously light-adapted. To reduce the probability of adventitious errors of estimation the experiments were made at a temperature high enough to bring the time for practically complete dark-adaptation down to a little over one hour. At about 12°, the time required may be as long as three to four hours; at 20.5°—22°, the temperature prevailing throughout the present experiments, the time is less than half of this. Since light-adaptation, *as measured by the power to influence geotropic creeping on a vertical surface,* is effectively complete within about five minutes, as a rule, with high intensities, the experiments must, for simplicity, take the form of establishing curves of light-adaptation as affected by known preceding periods in darkness. From such data it is required to deduce the ideal photic excitability at the last moment in darkness, before the exposure has begun, or in other words the kinetics of the "dark" reaction.

The simplest assumptions which serve to explain the data are the following. We suppose that the excitation at any moment is proportional to $-\frac{ds}{dt}$. The intercepts on the $\log_{10} \tan \beta$ axis at zero time of exposure to the light give measures of $-\frac{ds}{dt}$ at the first instant of exposure. The ratio $\tan \beta_0 / \tan \beta_t$, labelled R, where β_0 is the intercept angle for fully dark-adapted slugs, permits one to manipulate the resulting equations with somewhat greater ease. At different levels of dark-adaptation the formation of sensitive material must be taken into account. Various simpler hypotheses do not agree with the facts so far obtained, and the "dark" reaction in question has to be taken as a second order reaction with positive autocatalysis. The differential formula for such a process is:

$$\frac{ds}{dt} = [\, K' + K_2 \; (S_0 — x) \,] x^2,$$

where S_0 = concentration of S at complete dark-adaptation, x = concentration of P or A, and K' and K_2 are velocity constants. It is usually sufficient to neglect the constant K', which in most autocatalytic systems appears to be very small. The photolysis of S requires $-\frac{ds}{dt} = K_1 \; (S_0 — x)$. When both processes go on together,

$$-\frac{ds}{ds} = K_1 S_0 - K_1 x - K_2 S_0 x^2 + K_2 x^3.$$

The integral equation to which this gives rise is superficially formidable, but a variety of simple consequences at once derive from it and are found to correspond with directly ascertained experimental data:

$$-t = \frac{1}{K_1 - K_2 S_0^2} \left\{ \frac{1}{2} \log \frac{K_1 S_t^2}{S_0^2 \, (K_1 - K_2 S_0^2 + 2K_2 S_0 S_t - 2K S_t^2)} + \frac{K_2 S_0}{\sqrt{K_1 K_2}} \tan h^{-1} \frac{K_2 (S_0 - S_t)}{\sqrt{K_1 K_2}} \right\} .$$

For example, the second derivative of this expression gives the way in which the ratio R should change with time of dark-adaptation; it should be found that $1/R$, plotted against time, is a sigmoid curve with an inflection point occurring when the dark reaction is just a little less than 0.67 completed. The experimental value is 0.64 (Figures 44 *a* and *b*). The rate of light-adaptation at different levels of dark-adaptation is given by

$$-\frac{d^2 s}{dt \; dx} = K_1 - 2K_2 S_0 \, x + 3K \, x^2,$$

and this must pass through a minimum value at a period of dark-adaptation corresponding to the point at which the curve of $1/R$ is inflected (Figure 45). This is true experimentally. Moreover, for early stages of light-adaptation, at any level of dark-adaptation, as in the present measurements, log tan β (i.e., log S) should be almost linearly related to $-t$, since K_2 is known to be much less than K_1. Other deductions might be mentioned, but enough has been said to indicate the way in which the results of compounding tropistic effects can be employed, in this case, as a test of the tropism concept and for the development of its utility.

VIII

A major aspect of the theory of animal conduct is concerned with the initiation and the control of directed movements. If we regard the problem of "motivation" as having to do with the "springs of action," it may be supposed that the elucidation of the mechanism of the directed movements of even the most complex organisms will be materially furthered through the understanding of these aspects of behavior in forms more easily subjected to experimental control. The tropism doctrine successfully describes the course of one general class of such directed movements. It does not necessarily pretend to describe to what extent the movements of an organism in the complex field of a fluctuating environment will be conspicuously determined at a given moment by any specified type of tropistic response. For phototropic movements and for geotropic orientations under controllable conditions the relevant descriptions may be made precise, are physically interpretable, and are successfully employed for the prediction of movements in compound fields of excitation involving both photic and gravitational stimulations, even when time also enters as a significant variable. The theory of photic orientation, traditionally based upon rectilinear loci of progression, is moreover easily extended to deal with curvilinear paths, such as result from the concurrent operation of postural reflexes during orientation or from the effects of non-parallel rays of light. The fact that the resolution of certain of these questions was begun on the basis of experimental results obtained with mammals, later to be successfully employed for the analysis of corresponding modes of behavior among invertebrates, puts a somewhat different emphasis upon the tropism doctrine from that possible several years ago.

The tropism concept does not attempt to deal with the type of change in behavior which is evident when an organism may be moving about at one moment, totally quiescent at another, with external circumstances to all appearances identical. Nor is it concerned directly with such things as the experimental or the "spontaneous" reversal of sign in orientation, although such phenomena may be utilized; nor yet with the suppressions of orientation movements or responses such as may be brought about experimentally or may occur without deliberate intent on the part of the

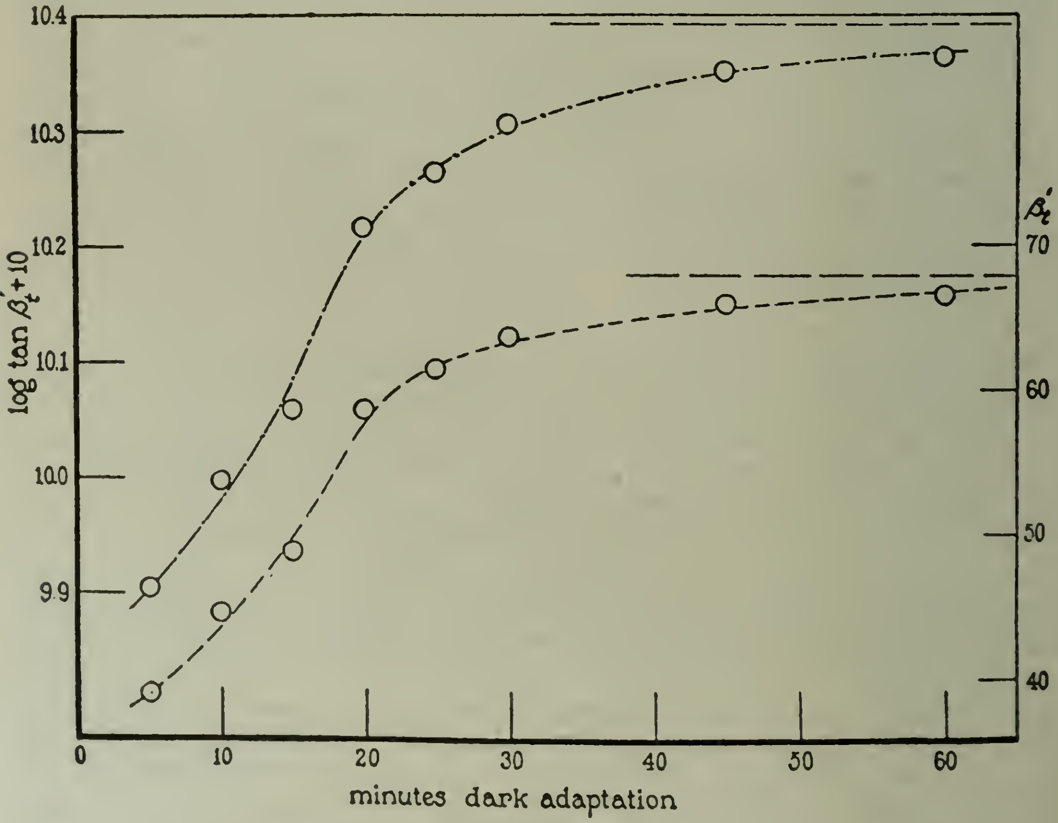


FIGURE 44*a*

Curves Showing the Increase of the Intercept Angle (β'_t), and of log tan β'_t, as a Function of Time in the Dark Subsequent to Prolonged Exposure to Light

(From *J. Gen. Physiol.*)

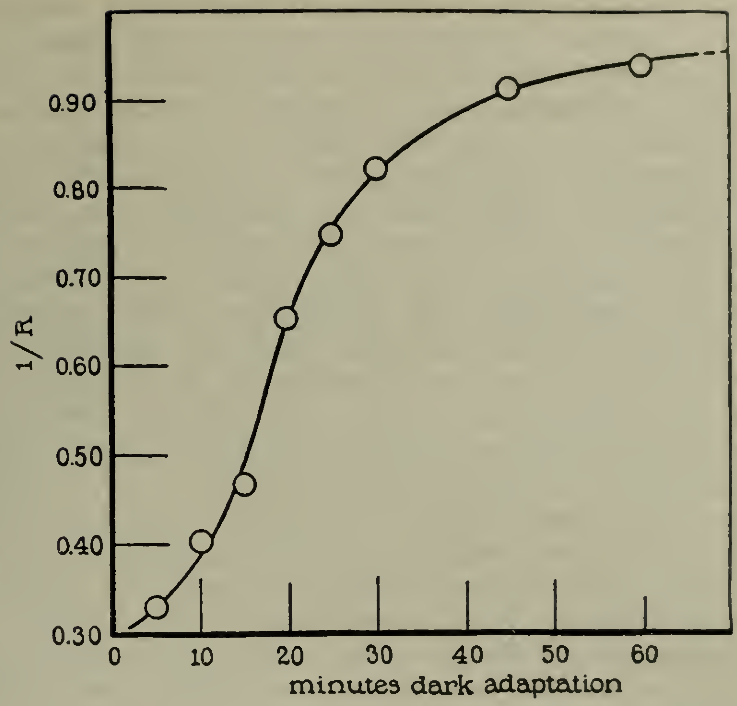


FIGURE 44*b*

The "dark" restitution of *S* through a second order reaction with positive autocatalysis would require that the photic excitability, $-\frac{ds}{dt}$, or its proportionate equivalent 1/*R*, should exhibit a maximum velocity of change when (S_t) has been brought to a little less than 2/3 (S_∞); 1/*R* should therefore pass through an inflection point when 1/*R* = a little less than 0.67—; the inflection is found at 1/*R* = 0.64. (From *J. Gen. Physiol.*)

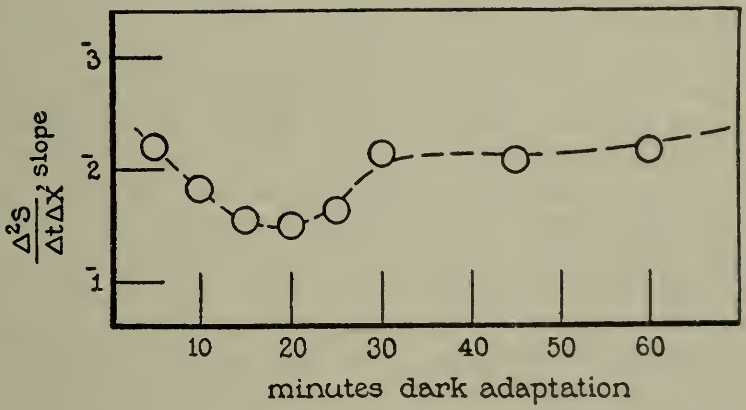


FIGURE 45

The second order "dark" formation of S requires that the rate of light-adaptation i.e., change of $-\frac{ds}{dt}$ should be a *minimum* when $-\frac{ds}{dt}$ is increasing most rapidly. The minimum found shows precise agreement with this requirement. (From *J. Gen. Physiol.*)

experimenter. The primary concern of the tropism doctrine is with the idea that in the case of phenomena to which it may properly be applied, orientation ceases when the excitations are equivalent upon the two sides of an organism structurally or dynamically symmetrical. From this assumption equations are derivable which describe the paths of movement under particular conditions, even when two contrasted sorts of stimulation produce a conflict which the organism must resolve. There would appear to be no fundamental distinction between the end result in such cases of this type as have been described for animals and those of dynamically comparable type well-known among plants. In the equations describing loci of orientation under such conditions there occur constants which may be assumed provisionally to measure, in the sense that they are determined by, properties of the adjustor mechanisms governing the manner in which the competitive effects of diverse excitations are resolved; these constants may then be studied from the standpoint of their possible experimental modification, with the hope that there may be approached in this manner some comprehension of the character of the mechanisms thus recognized and defined through their relations to controllable variables which may be employed to govern them. Investigations of this type appear likely to give a means for the study of properties of the central nervous system and its adjustor functions not otherwise open to measurement.

The history of the discussion of the quantitative investigation of oriented movements has suffered confusion through disagreements as to the precise steps in the process whereby an organism may become oriented, and in another respect by a rather unnecessary complication of descriptive terminology. It may be pointed out that to a large extent discussions in such terms are unprofitable and can in no sense affect the reality of precisely formulated relationships of orientation to excitation, although they may influence the use to which such formulations may be put as well as the analytical worth which it may be sought to attach to them. An interpretation sometimes conceived to be in profound opposition to the tropism concept is one which in a primitive form runs about as follows. At any moment the totality of molecular changes in an individual determines an internal state or condition, a "physiological state"; alteration in the energies of the environment constitutes, or may supply, a "stimulus"; neglecting for the moment the difficulties attending this definition of a stimulus, we may grant that as a result of the stimulus, movements are executed (creating a new "physiological state"); these movements are in certain respects governed as to their kinds, and indeed as to the occurrence of any movements at all, by the nature of the aforementioned "physiological state" and by those structural limitations which mechanically permit only certain sorts of movement to appear; they are, however, regarded as in essence movements executed without external constraint as to *direction*, and in large measure without constraint as to occurrence in time—that is, they are essentially *random* movements; by successive trials among these, the organism arrives at a position in which excitation is no longer experienced. One must also include here those movements appearing to arise as result of changes entirely within the organism, so far as can be told, which consequently appear as *spontaneous movements*; such movements seem usually regarded as free from constraints governing their incidence in time. In general, the so-called "random movements" tend to be related by writers to the natural or normal exploratory wanderings (so-called) of many organisms, the occurrence of which seems to be understood as accentuated or accelerated when the organism is "stimulated." A position of orientation is then assumed to be one such that excitation is reduced to a minimum. Everyone recognizes that under conditions not subjected to rigorous experimental control "aimless" and "spontaneous" or at least irregular motor activity is a feature of the conduct of animals. The profitable questions arising in this connection reduce practically to these: What rôle, if indeed any, do random movements play in tropistic orientation; and: Are "random movements" following excitation, spontaneous movements, and fluctuating physiological states such as may very properly be assumed to govern variability of response, really lawless? The successful formulation of tropistic movements in cases of the type which we have discussed provides an adequate answer to the first of these questions. It cannot possibly be assumed that the exhibition of a definitely oriented path in a compound field of excitation is governed by the assumption of a position such that excitation is reduced to a minimum or to zero. To the second question there has been as yet no generally satisfying answer.

We may with justice refer to the curious fact that in the vigorous discussion formerly prominent in connection with the notion of control of behavior through the determination of physiological states and adjustments on the basis of random explorations, few if any investigators considered it worth while to look into the quantitative aspects of the situation. The theory of random movements as customarily stated implies a hopelessly haphazard relation between stimulation and activation of motor organs, such as should vitiate any possibility of regularity and predictableness in the conduct of animals, and is to this extent contrary to experience under conditions where prediction is to be reasonably looked for. It is easily seen that the notion of "physiological states" is intimately connected, not only in an historical sense, with the random movement doctrine. The appeal to such states, in themselves not precisely characterized, easily leads to unfortunate mysticism. An adequate and typical illustration has been discussed by Parker (125) in an account of the way in which the activities of actinians are modified by feeding and withholding food.

We may consider three classes of phenomena in this general field, not with the thought that they exhaust the possibilities but rather that they may prove suggestive. There has been in general no treatment of the randomness of behavior such as would be required to permit a theory of the nature of this randomness and of the manner in which it may produce variability in conduct. At several points we have already discussed "variability" of response as shown under such conditions that units for

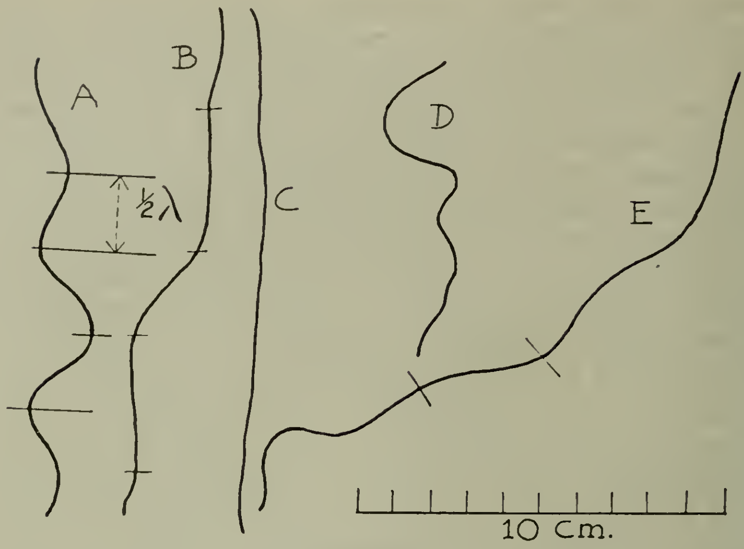


FIGURE 46*a*

Tracings of Actual Paths Obtained by Following a Moving Ant with a Soft Pencil

A, B, C, and *E* represent positive geotropic paths of *Lasius flavus; D* is a negatively geotropic path of *Aphaenogaster fulva;* slope of surface $= \alpha$. (Experiments by Dr. T. C. Barnes.)

A, $\alpha = 40$, illustrating the wide amplitude of the oscillations at a low inclination of the surface and the method of measuring the "wave-length" $1/2\ \lambda$.

B and *C,* the same ant at $\alpha = 70°$ and at $\alpha = 80°$.

D, Aphaenogaster with antennae removed, $\alpha = 65$. The antennae-less ant exhibits wider oscillations of path.

E, Lasius flavus, $\alpha = 30$. The ant was made to run rapidly by touching it at *X* with the pencil; the oscillations are "smoothed out" although the inclination is less than for trail *A*. The "θ" angle, however, is low.

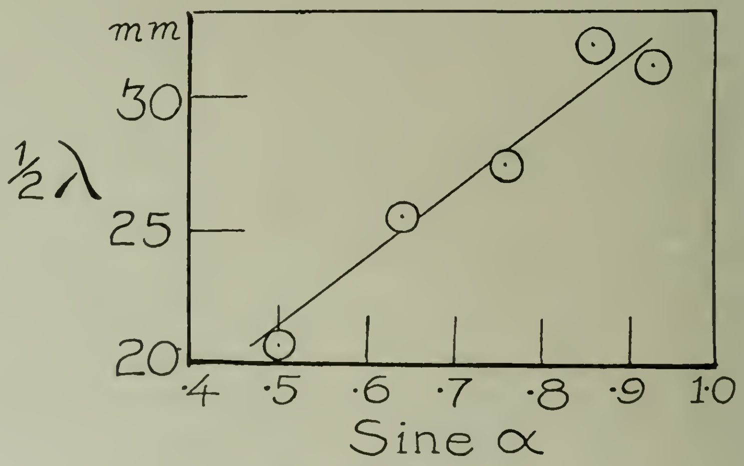


FIGURE 46*b*

The Relation between the Wave-Length ($1/2\ \lambda$) of the Oscillations in the Path of a Moving Ant and the Sine of the Angle of Inclination of the Plane

(*Vide* Figure 46*a* for the method of measuring $1/2\ \lambda$.) The increase of wave-length, paralleled by an increase of distance traversed between turns, measures the decreasing sinuosity of the path.

its measurements can be obtained. A further illustration is not out of place; this is taken from observations by Dr. T. C. Barnes upon the creeping of ants. When an ant such as *Aphaenogaster fulva* creeps upon a horizontal surface, its trail shows more or less definitely rhythmic alterations of direction; these constitute the well-known "exploratory" movements, seen sometimes with less regularity in other forms (Cornetz, 1914). The problem is: Are the movements which effect change of direction really lawless? This may be decided through investigations of the paths made by *Aphaenogaster* upon a surface which is inclined to the horizontal by different amounts in successive trials. This ant, as is likewise true of several other forms (Barnes, 1929), shows geotropic orientation. When this is examined in a long series of trials it is found that the "wave-length" of the path, that is, the distance between turnings to right or to left, which in the case of progression upon an inclined surface occur alternately, is in fact a definite function of the slope of the surface. Not only so, but the distance travelled (along the curved path) is similarly a function of the slope of the supporting plane. These points are illustrated in Figures 46 *a* and *b*. The appearance of random or intelligently exploratory movements here turns out to be in fact due to some mechanism easily controlled, and the elimination of random movements, leading to an increased straightness of the path pursued, proceeds according to a plan which we have already examined in the case of experiments with rats and is a simple function of the exciting condition.

The manner in which behavior in a field of excitation might be governed or limited through the exhibition of spontaneous movements, which might be taken to imply a kind of "free will" control, can perhaps be studied best when, with all possible care, it has been sought to exclude extraneous stimulation. In general it is found, and notably with lower organisms, that movements then still occur. Such activities may be termed spontaneous movements, since they appear not to depend upon peripheral excitation, and in some instances may indeed occur whether external excitation is deliberately applied or not. If it should prove that basically such movements occur in a manner simulating the automatic spontaneity exhibited by a contracting heart or a pulsating jellyfish, it scarcely remains possible to regard such movements as lawless or in any proper sense "random." In the experiments to be cited, which are due to Mr. T. J. B. Stier, it was sought to simplify conditions of observation by employing very young mice. In such animals it is possible to control the internal temperature of the organism by suitable change of that of the environment, the difference between the two being measured with a thermopile. When the young mice are placed in an observation chamber submerged in a thermostat, they are at first unable to creep, but do show spontaneous righting movements (attempts to assume erect posture) when they have fallen over to one side. The frequency with which such righting movements occur is a definite function of the temperature of the organism. In a particular genetic strain ("piebald") of these young mice it was found that a difference of temperature of 10° made a difference of two- to threefold in the frequency of occurrence of righting movements. The temperature characteristic $\mu = 23{,}000$ best describing the data agrees satisfactorily with that for the frequency of breathing movements in this strain; at 20°C., moreover, in each case, there occurs a definite "critical temperature" at which irregularities in the curve are quite evident.

If one attempts to examine more elaborately the control of the exhibition of spontaneous movements by temperature, it is not sufficient to proceed according to what might at first sight be thought the simplest technique. If one constructs a delicately adjusted diaphragm connected with devices permitting registration of movement, and places a young mouse upon the diaphragm, it quickly becomes apparent that there is no regularity visible in the incidence of movements of the legs or of the head; the reason for this unquestionably lies in the fact that the position of the organism is continually altered with reference to the substratum as soon as it begins to move. Greater refinement of manipulation is necessary. This has been achieved by supporting the very young mouse in a special harness astride a light celluloid frame so that only the feet are permitted to touch the diaphragm of the recording apparatus. The animal and the diaphragm with its connecting levers are enclosed within a chamber immersed in a thermostat, the temperature of which may be regulated; through the chamber filtered "outside" air, properly moistened and brought to the temperature of the thermostat, is passed at a constant rate of flow. The chamber is entirely dark. Under these conditions it is found that a pronounced regularity is apparent in the rhythmic recurrence of outbursts of activity and in the measurable durations of the successive alternating periods of activity and of quiescence. A typical record is exhibited in Figure 47. The frequency of incidence of periods of active movement is again found to be describable as a straight-line function of

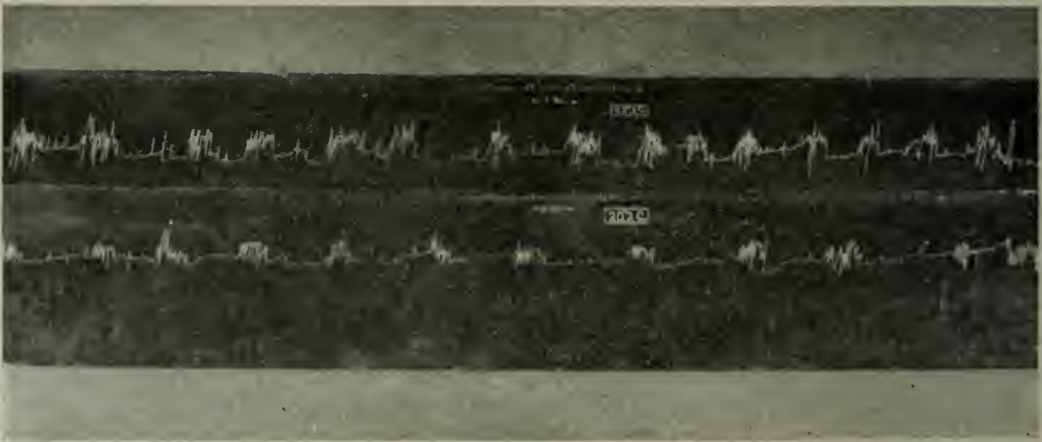


FIGURE 47

Records of Spontaneous Activity of a Young Mouse (Two Days Old) at Two Temperatures of the Body, Showing Regularity in Occurrence of Periods of Active Movement

(See text.)

$e^{\frac{-\mu}{RT}}$, where T is the absolute temperature, and μ has the value of 25,000±, the upper critical temperature being for this strain at 25°. The frequency with which the "impulse to move" (if we may use such terminology) becomes effective obeys in a predictable manner an elementary law of chemical dynamics. At any given temperature, moreover, the mean durations of intervals of active movement and of quiescence are statistically constant within narrow limits; for example, at a body temperature of 21.0° the periods of activity average 0.524 min., with a mean probable error = 4.7 per cent of the mean duration; the periods of quiescence average 1.73 min., with a probable error (larger than for the activity periods) = 11 per cent of the mean.

The behavior of older individuals shows considerable similarity as regards the periodic incidents of explosive outbursts of active movement. With advancing age these periods of activity become much longer, as do also the intervals of quiescence. They continue, however, to show certain definite although intricate regularities; the precise details differ according to the strain of mice used for the observation.

There is evidence to warrant the idea that the occurrence of these manifestations of intermittent activity are definitely connected with the behavior of the alimentary tract. Loeb (18) recognized the desirability of securing a theory of the "restless" movements of lower animals, and, noting the unceasing movements of the annelid *Nereis* after excision of its supraoesophageal ganglion, he suggested that the processes going on in the alimentary tract could then no longer influence the animal's behavior, since nervous connection with the central nervous system was interrupted. He conceived it possible that stimuli conducted through the nerves supplying the intestine might modify the alternation of activity and rest usually seen in the intact animal. This general view has since been substantiated and amplified, especially through the work of Boldyreff (1916) and Carlson (1916). Particularly from the results of Boldyreff it may be seen that there occur synchronous periodic fluctuations in the activity of the musculature of the stomach and small intestine, and of the glandular activity of the pancreas and other digestive glands. He was able to show that the "spontaneous activity" of an animal is definitely associated with vigorous contractions of the empty stomach. Rather extensive observations were made by Miss Tomi Wada (1922) on human subjects, who recorded at the same time hunger contractions and movements of the body during sleep. She found that the movements of a sleeping person occur only during the most intensive movements of the empty stomach, the so-called "hunger contractions" (Cannon and Washburn, 1912; Carlson, 1916). Such contractions of the quiet stomach can be initiated by increasing the intragastric pressure, with movement of the animal resulting. Food delays the usual occurrence of hunger contractions and also delays the occurrence of the next period of spontaneous activity. Drugs which relax smooth muscle, so-called spasm-inhibiting drugs such as papaverin, offset the periodic occurrence of hunger contractions and also lessen the bodily movements of the organism. Richter (129) found that periods of activity recur (in rats) about every hour when the animal is fed once a day, and last for about one-half hour; whereas when the animal is allowed to feed freely, periods of activity occur about every two hours and last for about one hour.

During the recurrent periods of increased tonic and contractive activity of an almost empty stomach, there is apparent a certain type of humoral control, as evidenced by the behavior of transplanted stomach cultures and of transplanted bits of pancreas (Ivy, 1925-28). It may therefore be assumed that the primary initiation of the periodic motor discharges arises in the alimentary tract. The relationship of the underlying changes to temperature suggests the probability of a fundamental chemical control of what we started by considering as perhaps a purely sporadic, random, and possibly lawless aspect of conduct.

The contrast between quiescence and movement is seen in an extreme form, nevertheless of widespread occurrence, in the phenomenon of tonic immobility ("death feigning"; cf. 47, 61, 92).

Reflex assumption of a motionless state, often supposed to simulate death and hence commonly referred to as "death feigning," is well known to be exhibited by a great variety of animals. Many unsuitable names have been applied to this condition. We shall speak of it as *tonic immobility*. In its typical manifestations this tonic immobility is to be distinguished sharply from mere cessation of neuromuscular activities; those cases with which we shall deal are characterized by a continuous (but, in at least some instances, plastic) tonic contracture of the muscles of the body and especially of the appendages. The phenomenon is developed to a notable degree among arthropods. A number of writers have discussed its natural history and to some extent its structural basis. Rabaud (1919) has recently given an extensive account of reflex immobilization among arthropods, and Mangold (1920) has prepared a digest of most of the older observations upon the general subject (cf. also Andova, 1929). The contributions have usually dwelt at length upon the supposed utility or otherwise of "death feigning," and Mangold, in particular, in connection with suggestions by a number of earlier writers, has attempted to erect a terminological scheme whereby "simulation of death" may be brought into relation with hypnosis, catalepsis, and analogous states characterized by the cessation of spontaneous movements (cf. also Piéron, 1913).

Singularly little attention has been given to the physiological analysis of tonic immobility. This is the more surprising because the phenomena of immobilization would seem to give opportunity, as I shall attempt to show, for convenient inquiry as to the sorts of forces and processes operative in a central nervous functional activity having sensible duration. These we have tested by means of experiments with the terrestrial isopod ("sowbug") *Oniscus asellus* Linn. With various arthropods, and in an amphibian, it has been noticed that the duration of an act of reflex immobility is influenced by the temperature. Such influence was found by Holmes (1906), the Severins (1911), Polimanti (1911), Szymanski (1912), Löhner (1919), and Grassé (1922). No attempt was made by any of these observers to determine the actual form of the relation to temperature.

The length of an act of immobility initiated by appropriate stimulation of an insect or crustacean is in reality determined by the duration of a sustained condition in the central nervous system. It is easily shown that in a variety of hexapods and isopods, with which the experiments were concerned, the animal gives other well-defined reflexes without emergence from the specific immobile state. With *Ranatra,* in particular, several striking proofs are obtained of the reciprocal innervation of the appendages and of the orientations of their central connections.

Nevertheless, as Rabaud has insisted, in each organism there are certain peripheral loci, stimulation of which at once arouses an immobile specimen at the will of the experimenter.

The duration of successive acts of immobility is rhythmic. In isopods and lizards at least, and probably also in such mammals as the guinea pig, the rhythm does not exist as a general metabolic rhythm on which the observations are, so to speak, superimposed. It is initiated by the first stimulation of a series. Once initiated, however, a cycle rises to a maximum of possible duration, then falls to practically zero duration, independently of further excitations. This is proved by results of tests in which various known intervals elapse between acts of immobilization, and by tests in which the animal is aroused to normal activity before its spontaneous "emergence," then re-immobilized.

The stimulus evoking tonic reflex immobility thus serves apparently to release a system of events, possibly synaptic in location, which determines at any moment the duration of a then initiated control of motor elements by suitable stimulation. The nature of this fundamental cycle, of quite appreciable duration, is deducible from the form of the cycle of successive immobilizations.

If suitably stimulated following spontaneous emergence from induced immobility, the isopod again becomes immobile, but for a longer time; the duration of later acts of immobility decreases. Continuously repeated tests of this sort show a periodic rise to a certain maximum duration. The amount of this maximum duration is determined chiefly by the temperature. At 6° the maximum duration is 270 sec.; at 29°, 23 sec. The cycle of immobility is not the reflection of a general metabolic cycle but is determined by a central nervous process initiated by the first stimulations of the series of trials.

The relation of temperature to the processes underlying the duration of reflex immobility involves comparing the durations at different temperatures but at *corresponding stages of the cycle.* Between 5° and 16° the maxima lie upon one exponential curve with respect to temperature; above 16°, and continuing to 30°, the maxima lie upon an intersecting exponential curve of lesser slope. The exponential relation of duration to temperature makes it likely that the duration is determined by an amount of substance produced in a chemical system, rather than that the maxima in the duration curves are to be considered points of minimum "speeds of emergence" from the immobile state. The curves are in fact fitted by Arrhenius' equation, $1/\text{time} = e^{\frac{-\mu}{RT}} + \text{const.}$, with $\mu = 24{,}000$ for the range 5° – 16°, $\mu = 9{,}200$ for 16° – 30°. These values of μ are within the range characteristic for chemical processes. More precise analysis of these relations is possible upon the basis of detailed consideration of the kinetics of the assumed underlying processes.

It appears, then, that the duration of an act of reflex immobility may be determined by a condition similar to synaptic transmissivity, continuously maintained for the duration of the act, and itself proportional to the amount of a substance formed by and broken down in two catenary processes having different temperature coefficients.

The durations of successive periods of induced tonic immobility in the lizard *Anolis carolinensis* were also examined as a function of temperature. The durations of the immobile periods were found to vary rhythmically in most cases (92). The reciprocal of the duration of the rhythm, i.e., the rate of change of the process underlying the rhythms, when plotted as a function of temperature according to the Arrhenius equation, shows distributions of points in two straight-line groups. One of these groups or bands of points extends throughout the entire temperature range with a temperature characteristic of approximately $\mu = 31{,}000$ calories, and the other covers the range of 20° to 35°C., with μ equal to approximately 9,000.

These results are interpreted by assuming the release, through reflex stimulation, of hormonal substances, one effective between 5° and 35°C. and the other effective between 20° and 35°C. These substances are assumed to act as selective inhibitors of impulses from so-called "higher centers," allowing impulses from tonic centers to pass to the muscles. In some experiments a progressive lengthening in successively induced periods of immobility was observed. The logarithm of the frequency of recovery when plotted against time in most of these cases (i. e., except for a few in which irregularities occurred) gave a linear function of negative slope which was substantially unaffected by temperature. In these cases it is assumed that a diffusion process is controlling the amount of available *A* substance.

Injections of small amounts of adrenalin above a threshold value are found to prolong the durations of tonic immobility of *Anolis,* by an amount which is a logarithmic function of the "dose." It is possible that internally secreted adrenalin, above a threshold amount, may be involved in the maintenance of tonic immobility.

IX

"Simplicity" of experimental conditions means control of experimental conditions. It does not necessarily mean that the organism must not be surrounded with apparatus; it does often imply that the undisturbed, intact organism must be studied. Simplification of conditions through control may make possible fruitful distinctions between phenomena of (1) *difference*, (2) *variation*, and (3) *variability* in the investigation of conduct, which unquestionably have been confused. This requires the biological uniformity of individuals, measurements from which are to be averaged. The essence of the notion of *variability* is found in the conception of variation (of response or performance) as a function of some controlling variable; otherwise, no proper *measure* of variability is obtainable. With suitably chosen conditions it is possible to separate the total variation over a range of fixed values of an independent variable governing the behavior in question into two parts, one of them modifiable according to the magnitudes of the governing intensity of excitation. In measurements of the geotropic orientation of young rats the variability of the measured orientations can be expressed in this way as a number which is independent of the size of the sample (within the limits used), and is characteristic for the stabilized strains employed in the experiments. In such cases it is no longer possible to speak of a lawless variation or unpredictableness of behavior.

When attention is given to the formulation of oriented movements, these are found to be machine-like in the sense that the conditions for stable position are easily interpreted in a physical way. The statement of such conditions involves a mathematical expression of functional dependence, which ceases to be arbitrary when its contained constants can be shown to correspond to a biological reality. In at least one case it has been shown that they remain "constants" when subjected to the purely biological manipulation involved in a breeding test. In other instances they provide clues for further analysis. On the basis of such formulations it has been possible to account quantitatively for the conduct of tested organisms in compound fields involving competition of excitations and change with time. It may be at least serviceably provocative to suggest that in this way a measure can be had of certain adjustor properties of the central nervous system. At least, it can be foreseen that through the difficult but attractive synthesis of more complex situations there is good likelihood of arriving at some measure of understanding of what animals do, and that the likelihood of this apears to be greater than that by "working downward" from the analysis of the most complex forms, problems, actions.

REFERENCES

The foregoing account has been restricted, deliberately, to the treatment of certain viewpoints not elsewhere stressed in quite the same manner. In a corresponding way the appended list of references makes no attempt at completeness. The critical enumeration of efforts to treat conduct as a department of physics would far transcend the scope of present intention. Particular phases of the subject, moreover, are discussed in other chapters of this book.

For aid in mathematical treatment the following will be found useful:

1. Arrhenius, S. Quantitative laws in biological chemistry. London, 1915. Pp. vii+163.
2. Daniels, F. Mathematical preparation for physical chemistry. New York, London, 1928. Pp. x+303.

3. Hitchcock, —, & Robinson, —. Differential equations. New York, 1923.
4. Michaelis, L. Einführung in die Mathematik. (3rd ed.) Berlin, 1927.
5. Thompson, S. P. Calculus made easy. London, 1919.
6. Walther, A. Einführung in die mathematische Behandlung naturwissenschaftlicher Fragen. Part I. Berlin, 1928. Pp. viii+220.

Sufficient introduction to the literature of animal conduct, with especial emphasis upon the analyzable behavior of invertebrates and lower vertebrates, is given in a number of works containing extensive bibliographies:

7. Claparède, E. Tierpsychologie. Handwörterbuch der Naturwissenschaften. Jena, 1913.
8. Hempelmann, F. Tierpsychologie vom Standpunkte des Biologen. Leipzig, 1926. Pp. vi+676.
9. Herter, K. Tastsinn, Stromungssinn und Temperatursinn der Tiere und die diesen Sinnen zugeordneten Reaktionen. Zool. Bausteine, Vol. 1. Berlin, 1925. Pp. iv+182.
10. Holmes, S. J. Studies in animal behavior. Boston, Toronto, 1916. Pp. 266.
11. Jennings, H. S. Behavior of lower organisms. New York, 1906.
12. Kafka, G. Einführung in die Tierpsychologie. Vol. 1. Leipzig, 1914. Pp. xii+593.
13. ————. Tierpsychologie. Handbuch der vergleichenden Psychologie. Part 1. Munich, 1922.
14. Köhler, O. Sinnephysiologie der Tiere. (Uebersichtsreferat.) *Jahresber. ges. Physiol.*, 1925 to date.
15. ————. Untersuchungsmethoden der allgemeinen Reizphysiologie und der Verhaltensforschung an Tieren. In: Peterfi, T. Methodik d. wissenschaftl. Biologie. Berlin, 1928. Vol. 2, pp. 845-925.
16. Kühn, A. Die Orientierung der Tiere im Raum. Jena, 1919. Pp. 71.
17. ————. Phototropismus und Photataxis der Tiere. Handbuch der normalen und pathologischen Physiologie. 1926. Vol. 12.
18. Loeb, J. Physiology of the brain. New York, London, 1900. Pp. x+309.
19. ————. Dynamics of living matter. New York, 1906. Pp. xi+233.
20. ————. Die Tropismen. In Winterstein's Handbuch der vergleichenden Physiologie. Vol. 4. 1913.
21. ————. The organism as a whole. New York, London, 1916. Pp. x+379.
22. ————. Tropisms, forced movements, and animal conduct. Philadelphia, London, 1918. Pp. 209.
23. Maxwell, S. S. Labyrinth and equilibrium. Philadelphia, London, 1923. Pp. 163.
24. Parker, G. H. The elementary nervous system. Philadelphia, 1919. Pp. 229.
25. Watson, J. B. Behavior. New York, 1914. Pp. xii+439.

Papers and books here listed have reference to specific points raised in the text; attention has been given chiefly to recent articles.

26. Andova, A. Thanatose des Grossen Rosskäfers *Geotrupes stercorarius* L. *Zsch. f. Morphol. u. Oekol.*, 1929, **13**, 722-744.
27. Arey, L. B., & Crozier, W. J. Homing habits of the pulmonate mollusk *Onchidium*. *Proc. Nat. Acad. Sci.*, 1918, **4**, 319-321.
28. ————. The sensory responses of *Chiton*. *J. Exper. Zool.*, 1919, **29**, 157-260.

29. ————. Natural history of *Onchidium*. *J. Exper. Zool.*, 1921, **22**, 443-502.

30. Baldi, E. Studi sulla fisiologia del sistema nervoso negli insetti, II. *J. Exper. Zool.*, 1922, **36**, 211-288.

31. Baldus, K. Untersuchungen zur Analyse der Zwangsbewegungen der Insekten. *Zsch. f. vergl. Physiol.*, 1927, **6**, 99-149.

32. Bancroft, F. W. Heliotropism, differential sensibility, and galvanotropism in Euglena. *J. Exper. Zool.*, 1913, **15**, 383-428.

33. Boysen-Jensen, P. Ueber die Leitung des phototropischen Reizes in Avenakeimphlanzen. *Ber. d. dtsch. bot. Gesell.*, 1910, **28**, 118.

34. ————. Ueber die Leitung des phototropischen Reizes in der Avenakoleoptile. *Ber. d. dtsch. bot. Gesell.*, 1913, **31**, 559.

35. Buddenbrock, W. v. Die Lichtkompassbewegung bei Insekten. *Sitzber. d. Heidelberg Akad. d. Wiss., Math.-naturwiss. Kl.* (*B*), 1917, **7**.

36. ————. Die vermutl. Losung d. Halterenfrage. *Sitzber. d. Heidelberg Akad. d. Wiss., Math.-naturwiss. Kl.* (*B*), 1917, **1**. Pp. 26.

37. Castle, E. S. Dark adaptation and the light-growth response of *Phycomyces*. *J. Gen. Physiol.*, 1928-29, **12**, 391-400.

38. Clark, L. B. Adaptation versus experience as an explanation of modification in certain types of behavior (circus movements in *Notonecta*). *J. Exper. Zool.*, 1928, **51**, 37-50.

39. Cole, W. H. Circus movements of *Limulus* and the tropism theory. *J. Gen. Physiol.*, 1922-23, **5**, 417-426.

40. ————. Geotropism and muscle tension in *Helix*. *J. Gen. Physiol.*, 1928, **8**, 253-263.

41. Crozier, W. J. The orientation of a holothurian by light. *Amer. J. Physiol.*, 1914, **36**, 8-20.

42. ————. The sensory reactions of *Holothuria surinamensis* Ludwig. *Zool. Jahrb., Abt. f. Zool. u. Physiol.*, 1915, **35**, 233-297.

43. ————. The behavior of holothurians in balanced illumination. *Amer. J. Physiol.*, 1917, **43**, 510-513.

44. ————. Assortive mating in a nudibranch, *Chromodoris zebra* Heilprin. *J. Exper. Zool.*, 1918, **27**, 247-292.

45. ————. On tactile responses of the de-eyed hamlet (*Epinephelus striatus*). *J. Comp. Neur.*, 1918, **29**, 163-175.

46. ————. On the control of the response to shading in the branchiae of *Chromodoris*. *J. Gen. Physiol.*, 1918-19, **1**, 585-591.

47. ————. Reflex immobility and the central nervous system. *Proc. Soc. Exper. Biol. & Med.*, 1923, **21**, 55-57.

48. ————. On stereotropism in *Tenebrio* larvae. *J. Gen. Physiol.*, 1923-24, **6**, 531-539.

49. ————. Wave length of light and photic inhibition of stereotropism in *Tenebrio* larvae. *J. Gen. Physiol.*, 1923-24, **6**, 647-652.

50. ————. On the critical thermal increment for the locomotion of a diplopod. *J. Gen. Physiol.*, 1924-25, **7**, 123-136.

51. ————. Biological researches relating to the nervous system. *Carnegie Instit. Yrbk.*, 1924, No. 23, 227-229.

52. ————. On biological oxidations as function of temperature. *J. Gen. Physiol.*, 1924-25, **7**, 189-216.

53. ————. On the possibility of identifying chemical processes in living matter. *Proc. Nat. Acad. Sci.*, 1924, **10**, 461-464.

54. ————. Galvanotropism and "reversal of inhibition" by strychnine. *J. Gen. Physiol.*, 1926-27, **10**, 395-406.

55. ————. The orientation of animals by opposed beams of light. *J. Gen. Physiol.,* 1928, **8**, 671-684.

56. ————. Tropisms. *J. Gen. Psychol.,* 1928, **1**, 213-238.

57. Crozier, W. J., & Arey, L. B. On the significance of the reaction to shading in *Chiton. Amer. J. Physiol.,* 1918, **46**, 487-492.

58. ————. The heliotropism of *Onchidium*: a problem in the analysis of animal conduct. *J. Gen. Physiol.,* 1919-20, **2**, 107-112.

59. ————. Sensory reactions of *Chromodoris zebra. J. Exper. Zool.,* 1919, **29**, 261-310.

60. Crozier, W. J., & Cole, W. H. Circus movements and heliotropism. *Anat. Rec.,* 1923, 362.

61. Crozier, W. J., & Federighi, H. On the character of central nervous processes. *Proc. Soc. Exper. Biol. & Med.,* 1923, **21**, 55-57.

62. ————. Phototropic circus movements of *Limax* as affected by temperature. *J. Gen. Physiol.,* 1924-25, **7**, 151-169.

63. ————. Suppression of phototropic circus movements of *Limax* by strychnine. *J. Gen. Physiol.,* 1924-25, **7**, 221-224.

64. ————. The phototropic mechanism in *Ranatra. J. Gen. Physiol.,* 1924-25, **7**, 217-221.

65. Crozier, W. J., & Libby, R. L. Temporary abolition of phototropism in *Limax* after feeding. *J. Gen. Physiol.,* 1924-25, **7**, 421-427.

66. Crozier, W. J., & Mangelsdorf, A. F. A note on the relative photo-sensory effect of polarized light. *J. Gen. Physiol.,* 1923-24, **6**, 703-709.

67. Crozier, W. J., & Moore, A. R. Homostrophic reflex and stereotropism in diplopods. *J. Gen. Physiol.,* 1922-23, **5**, 597-604.

68. Crozier, W. J., & Navez, A. E. Geotropic orientation of gastropods. *J. Gen. Psychol.,* 1930, **3**, 3-37.

69. Crozier, W. J., & Oxnard, T. T. Geotropic orientation of young mice. *J. Gen. Physiol.,* 1927-28, **11**, 141-146.

70. Crozier, W. J., & Pincus, G. Stereotropism in rats and mice. *J. Gen. Physiol.,* 1926-27, **10**, 195-203.

71. ————. The geotropic conduct of young rats. *J. Gen. Physiol.,* 1926-27, **10**, 257-269.

72. ————. Tropisms of mammals. *Proc. Nat. Acad. Sci.,* 1926, **12**, 612-616.

73. ————. Phototropism in young rats. *J. Gen. Physiol.,* 1926-27, **10**, 407-417.

74. ————. On the equilibration of geotropic and phototropic excitations in the rat. *J. Gen. Physiol.,* 1926-27, **10**, 419-424.

75. ————. Geotropic orientation of young rats. *J. Gen. Physiol.,* 1926-27, **10**, 519-524.

76. ————. On the geotropic orientation of young mammals. *J. Gen. Physiol.,* 1927-28, **11**, 789-802.

77. Crozier, W. J., & Stier, T. B. Critical thermal increments for rhythmic respiratory movements of insects. *J. Gen. Physiol.,* 1924-25, **7**, 429-447.

78. ————. The measurement of galvanotropic excitation. *J. Gen. Physiol.,* 1927-28, **11**, 283-288.

79. Crozier, W. J., & Wolf, E. On the place of photic adaptation. *J. Gen. Physiol.,* 1927-28, **11**, 289-295.

80. ————. Dark adaptation in *Agriolimax. J. Gen. Physiol.,* 1928-29, **12**, 83-109.

81. Dolley, W. L., Jr. Reactions to light in *Vanessa antiopa* with special reference to circus-movements. *J. Exper. Zool.,* 1916, **20**, 357-420.

82. Fries, E. F. B. Drug action in galvanotropic responses. *J. Gen. Physiol.,* 1927-28, **11**, 507-513.

83. Garrey, W. E. Light and the muscle tonus of insects. The heliotropic mechanism. *J. Gen. Physiol.,* 1918-19, **1**, 101-125.

84. Hartline, H. K. Influence of light of very low intensity on phototropic reactions of animals. *J. Gen. Physiol.,* 1923-24, **6**, 137-152.

85. Hecht, S. Sensory equilibrium and dark adaptation in *Mya arenaria. J. Gen. Physiol.,* 1918-19, **1**, 545-558.

86. ————. The photochemical nature of the photosensory process. *J. Gen. Physiol.,* 1919-20, **2**, 229-246.

87. ————. Intensity and the process of photoreception. *J. Gen. Physiol.,* 1919-20, **2**, 337-342.

88. ————. Sensory adaptation and the stationary state. *J. Gen. Physiol.,* 1922-23, **5**, 555-579.

89. ————. Intensity discrimination and the stationary state. *J. Gen. Physiol.,* 1923-24, **6**, 355-373.

90. ————. The visual discrimination of intensity and the Weber-Fechner law. *J. Gen. Physiol.,* 1924-25, **7**, 235-267.

91. ————. The kinetics of dark adaptation. *J. Gen. Physiol.,* 1926-27, **10**, 781-809.

92. Hoagland, H. Quantitative aspects of tonic immobility in vertebrates. *Proc. Nat. Acad. Sci.,* 1927, **13**, 838-843.

93. Holmes, S. J. The reactions of *Ranatra* to light. *J. Comp. Neur. & Psychol.,* 1905, **15**, 305-349.

94. Jennings, H. S. Contributions to the study of the behavior of lower organisms. *Publ. Carnegie Instit.,* 1905.

95. ————. Modifiability in behavior. I. Behavior of sea anemones. *J. Exper. Zool.,* 1905, **2**, 447.

96. ————. Modifiability in behavior. II. Factors determining direction and character of movement in the earthworm. *J. Exper. Zool.,* 1906, **3**, 435-455.

97. Just, E. Untersuchungen zur Frage der Gültigkeit des Resultantengesetzes. *Verhandl. dtsch. zool. Gesell.,* 1926.

98. Keeler, C. E. The geotropic reaction of rodless mice in light and in darkness. *J. Gen. Physiol.,* 1927-28, **11**, 361-368.

99. Kropp, B., & Crozier, W. J. Geotropic orientation in arthropods. III. The fiddler crab *Uca. J. Gen. Physiol.,* 1928-29, **12**, 111-122.

100. Loeb, J. Der Heliotropismus der Thiere. Würzburg, 1890. Pp. 118.

101. ————. Zur Theorie der physiologischen Light-und Schwerkraftwirkungen. *Pflüg. Arch. f. d. ges. Physiol.,* 1897, **54**, 81-107.

102. ————. Studies in general physiology. 2 vols. Chicago, 1905. Pp. x+782.

103. ————. The mechanistic conception of life. Chicago, 1912. Pp. 232.

104. ————. Forced movements, tropisms, and animal conduct. Philadelphia, London, 1918. Pp. 209.

105. Loeb, J., & Ewald, W. F. Ueber die Gültigkeit des Bunsen-Roscoeschen Gesetzes für die heliotropische Erscheining bei Tieren. *Zentbl. f. Physiol.,* 1914, **27**, 1165.

106. Mast, S. O. Light and the behavior of organisms. New York, 1911. Pp. 410.

107. ————. Photic orientation in insects with special reference to the drone-fly, *Eristalis tenax,* and the robber-fly, *Erax refibarbis. Amer. J. Physiol.,* 1923, **38**, 109-205.

108. ———. The process of orientation in the robber-fly, *Proctacanthus philadelphicus*. *Amer. J. Physiol.*, 1924, **68**, 262-279.

109. ———. Reactions to light in *Volvox*, with special reference to the process of orientation. *Zsch. f. vergl. Physiol.*, 1926, **4**, 637-658.

109*a*. Maxwell, J. C. Theory of heat. (10th ed.) London, 1908. Pp. 348.

110. Minnich, D. E. The photic reactions of the honey-bee, *Apismellifera* L. *J. Exper. Zool.*, 1919, **29**, 343-425.

111. Mitchell, W. H., Jr., & Crozier, W. J. Photic orientation by two point-sources of light. *J. Gen. Physiol.*, 1927-29, **11**, 563-583.

112. Moore, A. R. The mechanism of orientation in *Gonium*. *J. Exper. Zool.*, 1916, **21**, 431-432.

113. ———. Stereotropism as a function of neuromuscular organization. *J. Gen. Physiol.*, 1919-20, **2**, 319-324.

114. ———. Muscle tension and reflexes in the earthworm. *J. Gen. Physiol.*, 1922-23, **5**, 327-333.

115. ———. *Proc. Soc. Exper. Biol. & Med.*, 1923-24, **21**, 365.

116. Morgan, T. H. The theory of the gene. New Haven, 1926. Pp. xvi+343.

117. Morgulis, S. The movements of the earthworm: a study of a neglected factor. *J. Comp. Neur. & Psychol.*, 1910, **20**, 615-624.

118. Müller, A. Ueber Lichtreaktionen von Landasseln. *Zsch. vergl. Physiol.*, 1925, **3**, 113-144.

119. Müller, H. Die Lichtreaktionen von *Julus fallax* und *Polydesmus complanatus*. *Zool. Jahrb., Abt. f. allg. Zool. u. Physiol.*, 1924, **40**.

120. Navez, A. E. "Galvanotropism" of roots. *J. Gen. Physiol.*, 1926-27, **10**, 551-558.

121. Northrop, J. H., & Loeb, J. The photochemical basis of animal heliotropism. *J. Gen. Physiol.*, 1922-23, **5**, 581-597.

122. Oppenheimer, C., & Pincusson, L. Tabulae Biologicae. 4 vols. Berlin, 1925-27.

123. Paal, A. Ueber phototropische Reizleitungen. *Ber. d. dtsch. bot. Gesell.*, 1914, **32**, 499.

124. Parker, G. H. The directive influence of the sense of smell in the dogfish. *Bull. U. S. Bur. Fish.*, 1914, **33**, 63-68.

125. ———. The elementary nervous system. Philadelphia, 1920. Pp. 229.

126. Patten, B. M. A quantitative determination of orienting reaction of the blow fly larva (*Calliphora erythrocephala* Meigen). *J. Exper. Zool.*, 1914, **17**, 213.

127. ———. Photoreactions of partially blinded whip-tail scorpions. *J. Gen. Physiol.*, 1918-19, **1**, 435-458.

128. Pincus, G. Geotropic creeping of young rats. *J. Gen. Physiol.*, 1926-27, **10**, 525-532.

129. Richter, C. P. Animal behavior and internal drives. *Quar. Rev. Biol.*, 1927, **2**, 307-343.

129*a*. Rose, M. La question des tropismes. Paris, 1929. Pp. 469.

130. Stark, P. Studien über traumatotrope und haptotrope Reizleitungsvorgänge mit besonderer Berücksichtigung der Reizübertragung auf fremde Arten und Gattungen. *Jahrb. f. wiss. Bot.*, 1921, **60**, 67.

131. Stark, P., & Dreschel, O. Phototropische Reizleitungsvorgänge bei Unterbrechung des organischen Zusammenhangs. *Jahrb. f. wiss. Bot.*, 1922, **61**, 339.

132. Stier, T. J. B. Reversal of phototropism in *Diemyctylus viridescens*. *J. Gen. Physiol.*, 1925-26, **9**, 521-523.

133. Taliaferro, W. H. Reactions to light in *Planaria maculata*. *J. Exper. Zool.,* 1920, **31**, 59-116.

134. Torrey, H. B., & Hays, G. P. The role of random movements in the orientation of *Porcellio scaber* to light. *J. Anim. Behav.,* 1914, **4**, 110-120.

135. Weiss, P. Tierisches Verhalten als "System reaktion." *Biologia Generalis,* 1925, **1**.

136. Wheeler, W. M. The ant-colony as an organism. *J. Morph.,* 1911, **22**, 307-325.

137. ————. Social life among the insects. New York, 1923. Pp. vii+375.

138. ————. The social insects. New York, 1928. Pp. xviii+378.

139. Wolf, E. Physiologische Untersuchungen über das Umdrehen der Seesterne und Schlangensterne. *Zsch. f. vergl. Physiol.,* 1925, **3**.

140. ————. Geotropism of *Agriolimax*. *J. Gen. Physiol.,* 1926-27, **10**, 757-765.

141. Wolf, E., & Crozier, W. J. Orientation in compound fields of excitation; photic adaptation in phototropism. *J. Gen. Physiol.,* 1927-28, **11**, 7-24.

142. Yagi, N. Phototropism of *Dixippus morosus*. *J. Gen. Physiol.,* 1927-28, **11**, 297-300.

CHAPTER 3

THE MECHANISM OF REACTION

ALEXANDER FORBES
Harvard Medical School

In attacking the problem of the mechanism of reaction one naturally seeks to interpret the fundamental activities which constitute the basis of behavior in terms of physical law. The scientist should not be complacent in the assumption that all laws of matter, animate or inanimate, are known. But the hasty assumption that vital phenomena which do not appear to exemplify the known laws of matter are not subject to physical law, tends to confuse and obstruct progress. The law of the conservation of energy is no more violated by living matter than by other material systems. Those vital phenomena which baffle our attempts at explanation in terms of physical law should be regarded as complexities as yet unravelled, rather than as evidence that physical law does not apply to living things. As the marvels of physiology are elucidated by research, they show not so much new laws of life, breaking away from physics and chemistry, as examples of the ingenuity with which the physical laws and physical properties of matter have been utilized in the process of evolution.

NEUROMUSCULAR EVOLUTION

A study of the mechanism of reaction can perhaps best be approached through a consideration of the development and specialization of function in the neuromuscular system in the course of evolution. The protozoa are animals in which all the functions essential to life exist in a single cell. The amoeba, for example, is capable of digestion and assimilation of food, respiration, metabolism, excitation, conduction, motion, and reproduction. All these functions are performed by the single cell which constitutes the animal. In the present discussion we are concerned with the functions of excitability, conductivity, and motility, which, taken together and supplemented by secretion and certain other types of energy liberation, comprise the physical basis of reaction. Excitability may be defined as the capacity for responding with functional activity to changes in environment. Conductivity means the capacity to transmit a functionally active state from one part of the living substance to another. Motility is exemplified in the amoeba by the formation of pseudopodia and the flowing of cytoplasm from place to place; in ciliated protozoa and in ciliated epithelium of higher forms, by the beating of slender and flexible cilia; and in the major activities of all the animals higher in the scale than protozoa, by contraction of muscle.

The protozoa, having all necessary functions lodged in a single cell, are self-contained and complete in a very small compass, but their responses are stereotyped and only slightly adaptable to altered conditions. Their range of activities is thus extremely limited. Jennings (66) has shown that the ciliate protozoan *Paramecium* reacts to mechanical and chemical stimuli by the uniform response of reversing temporarily the beat of the cilia in such a manner that the animal backs a short distance, changes course, and proceeds again forward. This single response serves well enough to enable the paramecium to avoid obstacles and dangers in its path, but when the experimenter places a drop of strong acid behind the animal this same response causes it to back into the injurious substance and perish.

With the aggregation of cells into the large coordinated groups which constitute the bodies of higher animals, we find the needs of the organism served by the specialization of the functions of excitability, conductivity, and contractility in tissues differentiated for the purpose. Coordination is served by the mechanical arrangement of the body, as in the operation of the tube-feet of the starfish by means of the water-vascular system, or, more familiarly, in the action of the bones and joints of the vertebrate. Coordination is also effected by the production of special substances by glands and their distribution through the circulation. These two types of coordination will not be considered in detail here, for in this discussion the emphasis will be placed on coordination by the neuromuscular mechanism.

In a sense, excitability and conductivity cannot be dissociated. If excitation occurs at a single point and the response involves a finite volume of protoplasm, there is conduction. Conduction, in turn, requires the excitability of each successive region which becomes functionally active. Thus there cannot be excitation without conduction or conduction without excitation. On the other hand, we find in the fully developed nervous system a differential development of excitability and conductivity; that is, some structures are differentiated for the purpose of receiving stimuli, while others are differentiated for conduction.

The early steps in the evolution of the differentiated neuromuscular mechanism have been admirably sketched by Parker (95). His observations on sponges have led him to the conclusion that the first step toward a neuromuscular mechanism is the specialization of the contractile function in muscle. The sponge is an animal with primitive muscle-cells but with no trace of nerves. The function of the muscles is to close certain openings through which a current of water passes. They can be called into action by a variety of stimuli, but their response to these stimuli is always very sluggish. In the sea-anemone (a coelenterate) there are sensory cells in the integument, leading by slender nervous branches to the muscle-cells (Figure 1), marking the first stage in the evolution of a nervous system. This animal possesses a quicker and more effective type of responses. Yet the nervous processes of these sensory cells are very short, and, although the nerve net formed by them and the ganglionic cells renders nervous transmission possible from almost any part of the sea-anemone to almost any other part, the coordination is chiefly local. This is illustrated by the fact that a tentacle, when removed from the animal and held in sea water, gives almost exactly the same kind of response to the stimulus of food that it did

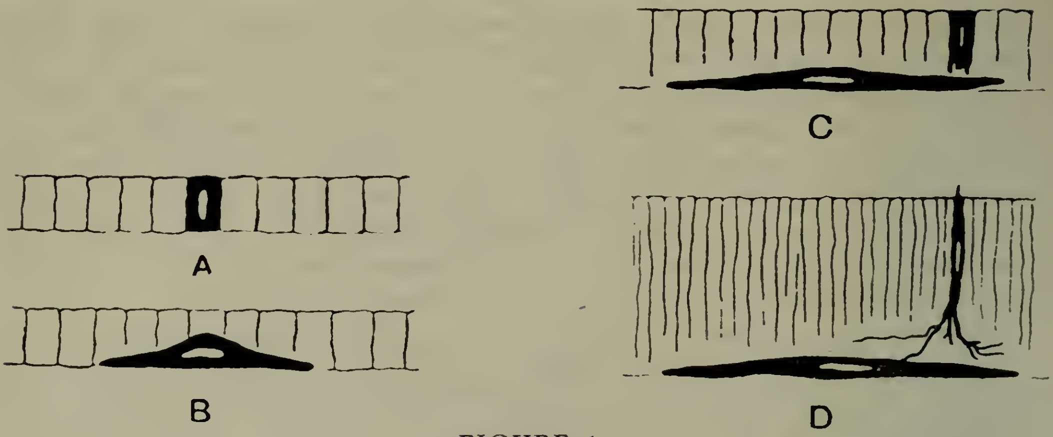


FIGURE 1

Diagram to Illustrate the Early Stages in the Differentiation of the Neuromuscular Mechanism

A, epithelial stage; *B,* differentiated muscle-cell at stage of sponge; *C,* partially differentiated nerve-cell in proximity to fully differentiated muscle-cell; *D,* nerve- and muscle-cell of coelenterate stage. (From: Parker, G. H. The origin and significance of the primitive nervous system. *Proc. Amer. Phil. Soc.,* 1911, **50,** 224.)

before removal from the animal. The nervous system, consisting of innumerable cells distributed throughout the integument, each connected only with its neighbors by short processes, is not characterized by any one dominant center. Summing up the structural features and the evidence from experiments on the animal's behavior Parker (94) concludes: "the nervous system in the sea-anemone is diffuse rather than centralized."

Parker (94) has summarized this evolutionary sequence as follows: "Muscle arose first and the simple *effectors*[1] thus produced were the first element of the neuromuscular mechanism. These effectors were directly stimulated and consequently slow in action. They afforded centers around which nervous tissues first differentiated in the form of sense organs or *receptors*[1] whose function it was to serve as triggers to initiate muscle action quickly. As these receptors became more highly developed, a third element, the central nervous organ, arose from the nervous elements between the receptor and the effector. This organ, the *adjustor,*[1] served as a means of conducting and modifying the sensory impulses on their way from the receptor to the effector and ultimately it also served as a storehouse for the nervous experience of the individual and as the seat of its intellectual life."

The next important step above the coelenterates appears in the worms. In the earthworm we find at the anterior end a centralized brain. This is connected with the skin by sensory nerve-fibers and with the muscles by motor nerve-fibers. Within the brain are the so-called association or *internuncial* neurons, nerve-cells which often form connecting links in the conduction path between the sensory and the motor nerve-fibers. Three salient steps have been made from the stage of the sea-anemone. The nerve-fibers or conducting processes of the nerve-cells have become very much longer, thus providing the mechanism for rapid coordination between different parts. Motor neurons have evolved, that is, nerve-cells whose function it is to conduct impulses to the muscles. A centralization of control has occurred; the central end of the sensory fibers, the association neurons, and the points of origin of the motor neurons being grouped together in a brain.

---

[1] Italics ours.

Correlated with this development we find in the behavior of the animal

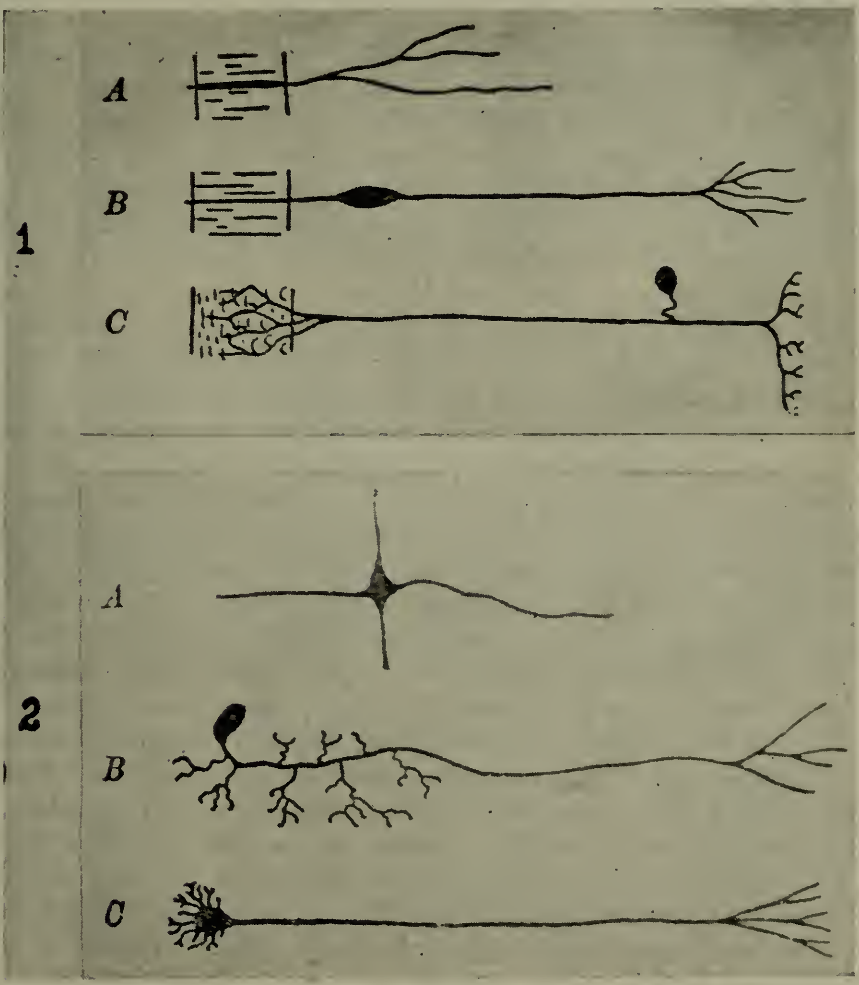


FIGURE 2

1. Stages in the Differentiation of Sense-Cells

A, sensory protoneuron from a coelenterate; *B*, sensory neuron from a mollusk; *C*, primary sensory neuron from a vertebrate. In each instance the peripheral end of the cell is toward the left, the central toward the right.

2. Stages in the Differentiation of Nerve-Cells

A, protoneuron from the nerve net of a coelenterate; *B*, motor neuron of an earthworm; *C*, primary motor neuron of a vertebrate. In *B* and *C* the receptive end of the neuron is toward the left, the discharging end toward the right. (From: Parker, G. H. The elementary nervous system. Philadelphia: Lippincott, 1919. Pp. 210, 211.)

an increased efficiency of coordination. "The earthworm responds to a large range of stimuli by appropriate and characteristic reactions, and its movements justify the conclusion that its reflex arcs, like those of higher animals, involve receptors, an adjustor, and effectors." Evidence is cited suggesting that the behavior of the earthworm may be modified by experience (94).

Parker points out that the three classes of nervous cells which occur in man are also represented in the earthworm. Thus the fundamental plan upon which the central nervous system of higher animals is organized is already laid down in an elementary form in the worm. Of course, the steps are many and important by which the nervous system is brought from the crude stage of the worm to that which makes possible the dominating intelligence of man, but perhaps the most important single element in the development is the increase in the number of association neurons. As Stiles (112) has put it: "With the addition of intermediate links between the receptor and the effector departments the reaction becomes less and less predictable . . . more and more . . . subject to reinforcement, suppression, or modification under the influence of changing circumstances."

Although the basic plan of neuromuscular organization, common to all higher forms, is already present in the earthworm, further development includes not only vast changes in the number of neurons and in the complexity of their arrangement, but also significant changes in the forms of the individual neurons. Parker gives a synopsis of the more salient changes, noting that in the evolving sensory neuron the cell-body migrates from the peripheral to the central end of the conducting path, and in the motor neuron the cell-body migrates toward the receptive end of the structure (Figure 2). Thus in each case the cell-body comes to occupy a central and therefore protected position in the body of the animal. Correlated with these changes is another of special significance; instead of the primitive nerve net, in which there is continuity between the conducting fibers of adjacent cells, the neurons make junction with one another at synapses at which there is believed to be contact but no protoplasmic continuity, and which are capable of conduction in only one direction. The typical vertebrate neuron, exemplified by the motor neurons of the spinal cord and most of the internuncial neurons of the central nervous system, consists of a nucleated cell-body with short, finely branching processes called dendrites, and a longer process, the axon, which conducts the impulses to the next neuron, or in the case of motor neurons, to the muscle. At the end of the axon are branches which, in the case of sensory and internuncial neurons, connect with the dendrites or cell-body of another neuron.

We arrive, then, at the general plan of neuromuscular organization found in the vertebrates, including man. The nervous system, embryologically of ectodermal origin, is made up entirely of neurons, histologically distinct units, linking the receptors and effectors. The simplest type of vertebrate reflex arc, providing for response to an external stimulus, is shown in Figure 3, which illustrates the arrangement of the neurons and

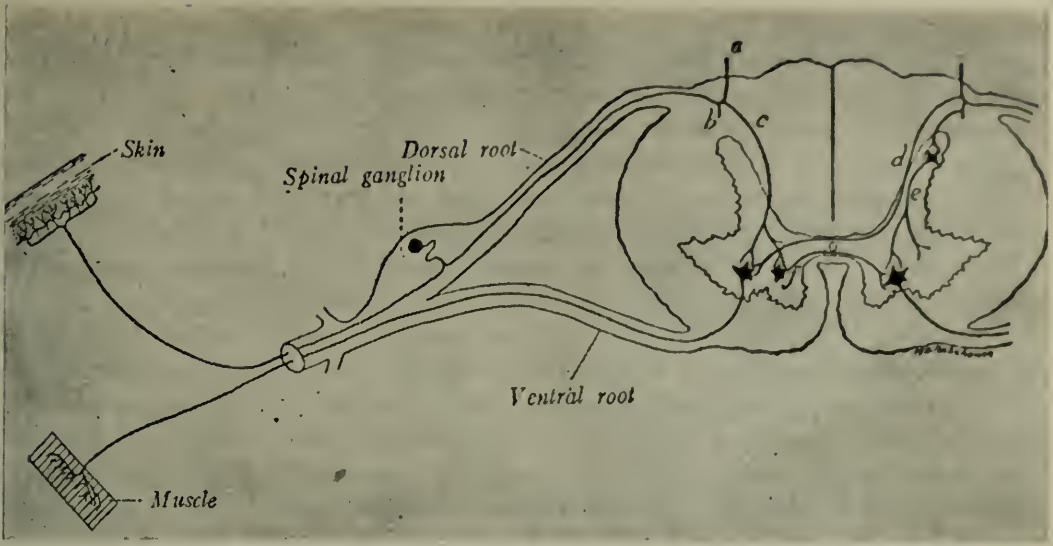


FIGURE 3

Diagrammatic Section through the Spinal Cord and a Spinal Nerve to Illustrate the Three Main Classes of Neurons *in situ*

A receptive neuron is shown in the dorsal root, a motor neuron in the ventral, and internuncial neurons in the cord itself (*e* and *o*). (From: Ranson, S. W. Anatomy of the nervous system. Philadelphia: Saunders, 1920.)

their central and peripheral connections. External stimuli in some instances act on free nerve-endings at the peripheral ends of the sensory neurons; in other instances these endings are connected with specialized sensory cells. In either case this portion of the reflex mechanism is termed the receptor. The working of the reflex arc is described by Parker (95) as follows:

"The receptors are connected by nerve-fibers with the central nervous organ or adjustor composed of the central ends of the sensory and the motor neurones and of the internuncial neurones. Here the impulses arriving from the receptors are directed toward the appropriate groups of muscles by which the animal may respond to the stimulus and, if the animal is highly organized, impressions are made upon the adjustor which, as memories, may become more or less permanent parts of the animal's nervous equipment. Finally the adjustors are connected by nerve-fibers with the third set of elements, the effectors, which as muscles, electric organs, glands, etc., enable the animal to react on the environment. Thus three physiological categories are to be distinquished which in the order of their sequence in action are sense organs or receptors, central nervous organs or adjustors, and muscles or other effectors."

Not only is this general plan common to all the vertebrates but so also are many of the anatomical details of the grouping and arrangement of the neurons. In man we find sensory neurons connecting all points of the integument and all the organs of special sense with the centers in the spinal cord and brain. The spinal cord is that portion of the central axis of the nervous system which lies outside of the cranium, or brain case. All the sensory neurons receiving impulses from the limbs, from the integument of the trunk, and from most of the internal viscera enter the spinal cord in segmentally arranged bundles known as dorsal roots.

The spinal cord is composed of white matter consisting of nerve-fibers, and gray matter consisting of nerve-cells and the terminal branches through which the neurons make connection with each other. The gray matter comprises the local nerve-centers or adjustors which exercise immediate control over the activity of the muscles of the limbs and trunk.

The brain is structurally a continuation of the spinal cord. Histologically it is similar in that it is made up of white matter, consisting of axons (conducting paths), and gray matter consisting of cell-bodies, dendrites, and terminal branches. In the lower vertebrates (e. g., fishes) the brain is a comparatively small enlargement at the anterior end of the spinal cord. As we ascend in the phylogenetic scale it becomes larger, till in man it is many times the size of the cord. This increase in relative size of the brain is clearly correlated with intelligence and complexity of behavior. The anatomy of the brain in man is extremely complex, and the millions of neurons of which it is composed vary greatly in size, shape, and arrangement, but they all possess the fundamental structures—dendrite, cell-body, and axon (or "neurite"), and they all fall into the three categories already named — sensory, internuncial, and motor.

Besides the central nervous system, comprising brain, spinal cord, and spinal nerves, vertebrates possess an autonomic nervous system consisting of outlying neurons whose cell-bodies are grouped in ganglia more or less remote from the brain and spinal cord, but connected with these central structures by neurons arising in the latter. This system is sometimes called the vegetative system, as its function is the regulation of the activities of the digestive tract, the smooth muscles of the circulatory system, glands, and other viscera not subject to voluntary control nor participating directly in the immediate reactions to the environment in the manner characteristic of the skeletal muscles.

There is also present in the digestive tract of vertebrates a nerve net resembling the primitive nervous system of the coelenterate. This nerve net, in contrast with the central nervous system, is devoid of synapses. Thus in the higher forms we find in conjunction with some smooth muscle-cells, resembling as they do the muscle-cells of sponges and coelenterates, a primitive nervous system which has survived almost unchanged through the course of evolution. Its extent, however, is insignificant compared with the synaptic nervous system which, in the higher forms, has assumed the dominant rôle.

It is important to note that, though all parts of the vertebrate body are coordinated by the nervous system, all conducting paths lead to or from the centers in the spinal cord or brain, never directly from organ to organ.

Effectors

The effectors at the disposal of the animal kingdom are of several kinds. There are nettling cells in coelenterates. There are cilia, slender processes which beat like the limb of a swimming animal. Cilia are found in protozoa, in which they serve the purpose of locomotion. They are found in the tentacles of many aquatic animals from coelenterates to polyzoa; they are found on epithelial surfaces of metazoa from coelenterates to mammals; in these cases they serve to move the surrounding fluid. There are chromatophores or pigment-bearing cells, capable of contracting and expanding and thus altering the appearance of the integument which contains them. These are found mostly in vertebrates. There are glands composed of cells which produce various substances, deriving the constituents from the body fluids or the blood stream. These include sweat glands, digestive glands, endocrine glands which pour their secretions into the blood stream, and glands producing such special secretions as the sepia whereby the squid makes a sort of aquatic smoke screen to confuse its adversaries. Glandular function is found in all classes of animals from protozoa to man. There are luminous organs in various forms beginning near the bottom of the animal scale and appearing as high in the scale as fishes. There are electric organs in certain fishes, specialized for the purpose of giving electric shocks to hostile creatures. Finally, there are muscles, which by contracting serve the purpose of motility. As we have seen, they are first specialized in sponges, and they constitute the most widespread type of effector in all classes of metazoa. In the vertebrates muscles are of two main types — smooth and striated. The smooth muscles are the more primitive type, resembling the muscles of the sponge and the sea-anemone; their contraction is slow and usually of comparatively sustained character. They are found chiefly in the walls of the blood-vessels and the hollow viscera. Striated muscle is specialized for rapid contraction. It moves the limbs, trunk, neck, jaws, tongue, larynx, and eyes. Because of its attachment to bony structures it is designated skeletal muscle. It serves the purpose of locomotion and all manner of other activities. It is normally classified as "voluntary." The heart muscle, though striated, is not subject to voluntary control and is physiologically intermediate in character between smooth muscle and skeletal muscle.

Of the above-named types of effector, the nettling organs are independent of the nervous system, being called into functional activity by the direct action of external stimuli. Cilia are usually independent effectors but in some cases are under nervous control. Chromatophores, glands, luminous organs, and muscles are in some cases under nervous control and in some cases not. Electric organs, which morphologically are akin to muscles, are always under nervous control.

Conduction

In considering the nature of the functional evolution which has enabled the neuromuscular mechanism to perform its intricate task, we shall deal first with conduction, for though, according to Parker, the specialization of muscle for the purpose of contraction preceded the specialization of nerve for conduction, protoplasmic conduction is the more fundamental property and is common to both muscle and nerve.

When the motor nerve of a vertebrate is stimulated, no visible change appears in the nerve itself, but the innervated muscle promptly responds by a vigorous contraction. An invisible disturbance has been propagated along the nerve-fibers at a much higher speed than the conducted disturbance in undifferentiated protoplasm, such as that of epithelial cells and protozoa. The nerve-fiber is highly specialized for the purpose of rapid and efficient conduction — up to 100 meters a second in mammalian motor nerves.

Excitation. Despite the obvious differences between nerve and muscle, both as to structure and function, the fundamental nature of their activity has so many important features common to both tissues that their differences seem to be more superficial than their resemblances. Both nerve and muscle may be excited by various classes of stimuli, mechanical, thermal, chemical, and electrical. There is reason to believe that the essential element in all forms of stimuli is an electric disturbance. The study of electrical excitation of nerve and muscle has thrown a great deal of light on the nature of their functional activity. The researches of Nernst (89, 90), Lapicque (70), Lucas (81), and Hill (56) have included quantitative measurements upon the relation between intensity and duration of a current sufficing to excite a given tissue. These researches have revealed the fact that when an electric current is passed through a portion of a nerve- or muscle-fiber a local change occurs which is limited to the region through which the current flows. If this local change does not attain a certain critical intensity, it produces no remote effects whatever; if it is made sufficiently intense, however, it will set up a disturbance which is conducted away from the point of stimulation over the entire length of the fiber. The "local excitatory process," as it was designated by Adrian and Lucas (12), persists for a time after the stimulating current has ceased to flow, and its rate of subsidence is characteristic of each tissue.

The prevailing interpretation of experiments on this subject is that the local excitatory process consists in a concentration of ions at some point in the tissue. Only when this concentration reaches a certain requisite value does a propagated disturbance sweep over the fiber. In spite of quantitative differences between nerve and muscle, especially as to time relations, the curves correlating the requisite intensity of the exciting current with its duration have the same shape for both tissues. The basic principle of excitation is clearly the same for both. A typical curve is shown in Figure 4. It will be noted that the curve ceases at a certain point to approach the base line, and is thereafter horizontal. This means that, up to a certain point, increasing the duration decreases the strength of current required to excite, but beyond that point further increase in duration is immaterial. This minimal strength of current is called the *rheobase.* For the sake of comparing excitation time in different tissues, Lapicque has introduced a constant, *chronaxie,* which serves to characterize the time element in this type of curve. A current of double the strength of the rheobase is arbitrarily chosen, and the least duration at which such a current will excite is determined. This duration is called the *chronaxie.* This measure of time varies from about 0.2σ in mammalian medullated nerve to 100σ in the case of smooth muscle.

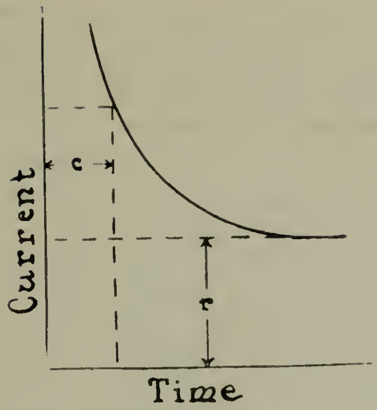


FIGURE 4

Typical Curve Correlating Duration with Strength of a Constant Current Which Barely Suffices to Excite

Abscissae, duration; ordinates, strength of current. *r*=rheobase. *c*=chronaxie. (Taken from observations by Lucas on sartorius muscle of toad: Lucas, K. The excitable substances of amphibian muscle. *J. Physiol.,* 1907-1908, **36**, 118.)

Propagated Disturbance. Even more striking evidence of the essential similarity in nerve and muscle is found in the many identical features revealed by the propagated disturbance which results when the local excitatory process becomes sufficiently intense. This disturbance has been known classically in the case of nerve as the "nerve-impulse," in the case of muscle as the "wave of excitation." The single designation *propagated disturbance* (Adrian and Lucas) is a noncommittal and fitting name for it in both tissues. Perhaps the most noteworthy feature of this functional response is the electric change by which we may follow its progress along the fibers. A suitable galvanometer or electrometer connected with a nerve or muscle reveals a wave of lowered electrical potential sweeping along the fiber, the active portion being negative with respect to the inactive regions in front of and behind the advancing wave. This effect, as observed in a recording galvanometer, is commonly called the "action current." Records from both nerve and muscle made with the capillary electrometer by Adrian (4) and from nerve with the cathode ray oscillograph by Gasser and Erlanger (48) and Bishop (16) show that the electric disturbance rises rapidly to a maximum and then dies away more slowly, the decline lasting about four or five times as long as the rise to maximum. The actual duration of the initial rise in mammalian motor nerve is about 0.3σ; in the most rapidly responding skeletal muscle it is about four times as long. Records of action currents in nerve and muscle, made with a string galvanometer, are shown in Figure 5.

A probable clue to the physical significance of the action potential is the "current of injury," which flows through a galvanometer connected at one terminal with the uninjured surface of a nerve or muscle, and at the other with the transversely cut end or otherwise injured portion of the tissue. The cut end or injured portion is negative with respect to the uninjured surface, just as during the passage of an impulse the active region is negative with respect to the inactive region. Furthermore, although it is impossible to infer the actual magnitude of the potential difference involved in either the action current or the current of injury, "since the circuit through the instrument is shunted to an unknown degree by idle tissue and fluid" (25), it is probably significant that the potential difference appears to be of the same order of magnitude in the two cases (presumably between 50 and 100 millivolts).

The most probable interpretation of all these observations is embodied in the *membrane theory,* which was developed for nerve by Ostwald (92) and Bernstein (15) and somewhat extended by Lillie (77, 78). This theory assumes a semipermeable membrane which normally maintains a difference in concentration of electrolytes, and hence a difference of potential, between the protoplasm within and the external surface of the fiber. The current of injury, obtained by transverse section of the fiber, constitutes probably its most direct measure, for the electrode at the cut end is applied as directly as possible to the exposed inside portion of the fiber. On the membrane theory the action current, which marks the passage of an impulse, is due to a transient increase in permeability resulting in a partial or complete depolarization moving progressively along the fiber. Even though there is no recording instrument there must be a local bio-

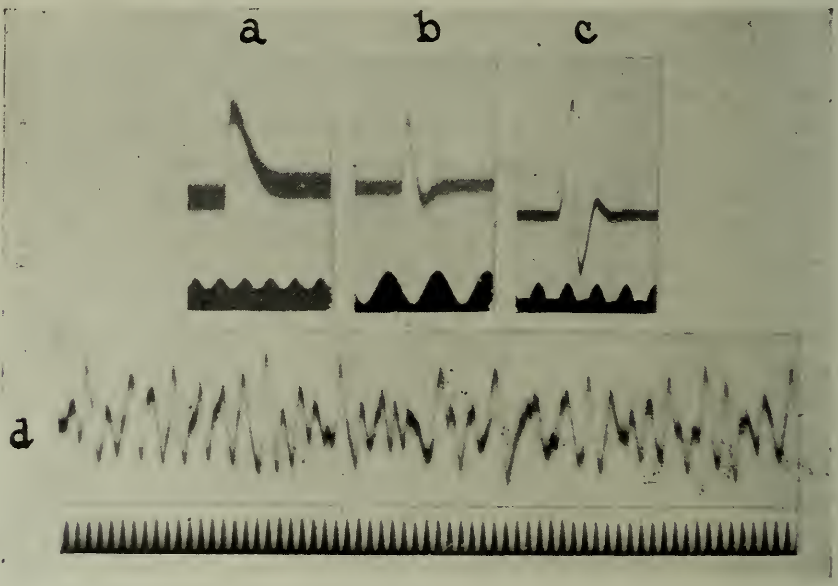


FIGURE 5

Records of Action Currents in Nerve and Skeletal Muscle Made with String Galvanometer

a, monophasic action current of popliteal nerve of cat; *b,* diphasic action current of peroneal nerve of cat; *c,* diphasic action current of gastrocnemius muscle of frog; *d,* action currents of human forearm flexor muscles during voluntary contraction, electrodes applied to the skin. In *a, b,* and *c,* the response was in each case evoked by a single stimulus (induction shock).

In all, time is shown below by shadow of tuning fork vibrating 100 d.v. per sec.

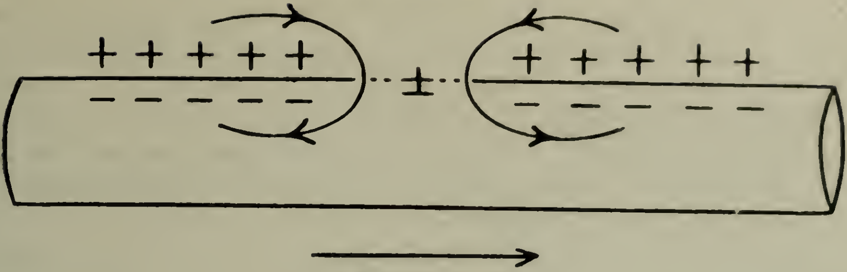


FIGURE 6

Diagram to Illustrate the Membrane Theory of Conduction in Nerve and Muscle

The arrow below indicates the direction of the propagated disturbance. The general course of the bio-electric currents is shown by the curved arrows.

electric current (78) in consequence of the action potential. This current is assumed to flow from the positive outer surface in front of the disturbance, along the outer sheath, and through the surrounding fluid to the active region, at that point inward through the depolarized membrane to the inside of the fiber, and back again along the inside (see Figure 6). It is further assumed that this current constitutes the stimulus which calls into action the next section of the nerve and thus causes the wave of activity to progress. The cessation of the action current is assumed to depend on the spontaneous restoration of the polarized state as the wave of activity passes on.

This theory is reinforced by the analogy of the behavior of an iron wire with a film of oxide on its surface immersed in nitric acid. The wave of electro-chemical activity which can be made to pass along the wire furnishes a striking counterpart of the nerve-impulse in certain respects (78). The close similarity of nerve and muscle with respect to both action currents and currents of injury leads to the conclusion that the mechanism of conduction must be fundamentally alike in the two tissues. In support of the membrane theory Lillie (76) has called attention to the correlation between the rate of rise of action current and velocity of conduction in various types of nerve and muscle. The hypothetical membrane cannot be identified with any particular part of the structure of the nerve-fiber. Bishop and Erlanger (16, 17) show features of the action potential which suggest the progress of chemical reactions, and in discussing the membrane theory they warn against any oversimplified and static explanation and regard both action and resting potentials as the resultants of balanced reactions.

Refractory Phase. Another feature common to nerve and muscle, and of the greatest importance, is the *refractory phase.* When a propagated disturbance is initiated it sweeps over the tissue, leaving it refractory to further stimulation for a brief interval of time (18). Following the absolute refractory phase when the tissue cannot respond, there is a period of recovery, the relative refractory phase, during which the threshold of excitation is abnormally high, and the size of the response which can be evoked is subnormal, both returning gradually to normal. Adrian has shown that in the case of a nerve immersed in a slightly acid fluid there is, following the relative refractory phase, an apparently *supernormal* phase in the recovery of excitability. He further showed that the excitability becomes no greater than would be found if the same tissue were in neutral fluid. The supernormality is only relative, due to the fact that in an acid medium the final resting stage is subnormal when compared with the condition of the tissue at neutrality (3).

Just as the action current is briefer in the case of nerve than in muscle and the velocity of conduction more rapid, so is the refractory phase briefer. There is not a strict proportionality but a very clear correlation as regards quickness of function in its various aspects (see Table 1).

TABLE 1

	Duration of action current		Refractory phase		Velocity of conduction (meters per sec.)
	Rising phase	Total	Absolute	Total (including relative)	
Medullated nerve (frog sciatic) 22° C.	0.3σ	2σ	2σ	8σ	30
Skeletal muscle (frog sartorius) 24° C.	3.0σ	20σ	3σ	11σ	2

The values, compiled from various sources, are only approximations. For total duration of both action current and refractory phase the approximations are very rough because the decline of the action current and the return of excitability to normal are both so gradual that precise determination of their duration is impossible.

All-or-None Law

It has long been known that when a single stimulus is applied directly to a muscle or to its motor nerve, the size of contraction which may be evoked is apparently graded according to the strength of stimulus from the threshold, below which the stimulus is ineffective, to a maximal value, above which further increase in the strength of stimulus causes no further increase in the size of contraction. Lucas (80), by using a very small muscle and a delicate means of recording the size of the contractile twitch, showed that the increase in contraction as the stimulus was gradually increased in strength occurred in a series of definite steps, and after each step was reached there was no further increase in the size of contraction until the next step. Pratt (99), using even more refined methods, brought out this same effect more strikingly. These observations can be interpreted only as meaning that each step represents the threshold of stimulation for a new fiber or group of fibers, and that the contractile effort of the individual fiber is the same whether the stimulus be strong or weak. Adrian (5), using the pore-electrode, showed that the action current of muscle likewise increases by a series of definite steps, the level being maintained constantly throughout the extent of each step, and that therefore the electric response, as well as the contractile effort of the individual muscle-fiber, obeys the *all-or-none law*.

In the case of nerve, Gotch (55) with a capillary electrometer showed that the action current attained a limiting maximal value, beyond which it did not increase with further increase in the strength of stimulus, and that records of submaximal action currents were of exactly the same shape as those of maximal action currents. He inferred from this that submaximal responses were simply those in which less than all the constituent fibers of the nerve participated and that the response of an individual fiber was always the same in form and amount.

Adrian in 1912 (1) attacked the question of the all-or-none character of the nerve-impulse by a new method. When a portion of a nerve is subjected to a narcotic, the impulse is blocked, provided the narcotic is sufficiently concentrated and applied for a long enough time. Apparently the duration of action of a narcotic necessary to abolish conduction depended on the length of the narcotized region. From this it was concluded that the impulse became progressively weaker as it traversed the narcotized portion of nerve. This is the classical conception of *"decrement."* Proceeding on the assumption of decremental conduction in a narcotized region, Adrian demonstrated that if a nerve-impulse is allowed to traverse a region of narcosis almost but not quite long enough to extinguish the impulse, it will, on emerging into an unnarcotized region, apparently regain its initial size. These and other observations (2) involving narcotization led to the conclusion that *the size of the nerve-impulse in the individual fiber is independent of the strength of stimulus,* i. e., *that it obeys the all-or-none law.*

Kato in 1924 (68) pointed out that the assumed decremental conduction in a narcotized region, which had thus been used to furnish evidence for the all-or-none law in normal nerve, constituted an exception to the all-or-none law itself. He and his colleagues showed that the dependence of the duration of narcosis upon the length of the narcotized region was limited to regions less than 7 mm. long. Beyond 7 mm. the length of the narcotized region was immaterial. In shorter lengths the observed relation is in part explainable through diffusion near the edge of the narcotizing chamber with a resulting gradient of concentration of the narcotic. These experimenters also showed that when the nerve-impulse enters a region of uniform narcosis the electric response falls to a reduced level, but instead of growing progressively smaller, as the decrement theory assumed, it remains constant in size as long as the degree of narcosis is constant. On emerging from the narcotized region to normal nerve the electric response at once regains its initial size. Davis, Forbes, Brunswick, and Hopkins (26), working independently, afforded complete confirmation of Kato's conclusion. These experiments, although disproving the premise of decrement in narcosis, upon which Adrian's chain of evidence rested, nevertheless confirmed his main conclusion and placed the all-or-none law on a firmer foundation than ever. Davis, Forbes, Brunswick, and Hopkins contended on theoretical grounds that there must be a brief transitional decrement as the impulse passes from a normal region to one of partial narcosis.

Cooper in 1926 (23) reported observations which appeared difficult to reconcile with Kato's view of decrementless conduction. The difficulty has been met by Rice and Davis (101), and we may definitely abandon the idea of progressive decrement in uniformly narcotized nerve, with the reservation that on theoretical grounds there must be some degree of transitional decrement, even if only of molecular dimensions, and the question of extent remains yet to be determined. Perhaps the strongest case for the all-or-none law in nerve is the reasoning which Adrian brought to bear on the problem in 1914 (2). Since the impulse is set in motion as soon as the local excitatory process reaches the requisite value "and the refractory state accompanying the disturbance will prevent any subsequent change in the local conditions from affecting its size in any way," the all-or-none relation is a necessary logical consequence.

The evidence that the impulse regains its full size on emerging from a narcotized into a normal region shows that *the energy of the impulse comes not from the stimulus but from the fiber.* This fact taken together with the refractory phase, denoting a temporary depletion of the source of energy, shows that the nerve-impulse is dynamically like an explosive reaction or the burning of a fuse, rather than like a sound wave or an electric current in an inert conductor. A disturbance of this class must be fundamentally independent of the strength or character of the stimulus which evokes it, provided this be adequate. It should be noted that the membrane theory of conduction harmonizes well with the all-or-none character of the response, for, if the essential feature of the impulse is a breaking-down of the polarized state of the membrane, it is evident that such a change cannot go beyond complete depolarization.

The all-or-none law of response is now well established for nerve and skeletal muscle. This law, frequently misunderstood, in reality states that *the release of energy (propagated disturbance) evoked by a single stimulus in a single functional unit of nerve or muscle is always as large as that functional unit is capable of producing at the moment when the response is evoked, no matter how strong the stimulus may be.* In other words, the size of the propagated disturbance is independent of the strength of the stimulus, provided this is adequate (38). The size of response may be varied by varying the condition of the tissue; for instance, the energy released in a muscle twitch may be varied by altering the load to which the contracting muscle is subjected. The size of response varies with the degree of recovery from a previous response. But under all these conditions the size of response is independent of the strength of stimulus.

The recent researches of Adrian and his collaborators (7, 8 , 9, 13, 14) have shown quite clearly the all-or-none character of sensory nerve-impulses and the fact that, while their magnitude is constant, their frequency is correlated with the strength of the stimulus applied to the sensory receptor. Thus sustained tension applied to a small strip of frog's muscle, containing a single receptor for muscle sense, set up in the afferent nerve a regular succession of impulses varying in frequency from 100 per second with the maximum load to 5 per second with a very weak pull. If the tension applied was constant, the frequency of the nerve-impulses declined gradually during its application.

METABOLISM IN NERVE

It has long been known that both nerve and muscle require oxygen in order that they remain functionally active, and that muscle gives off carbon dioxide as a result of its contractile activity. Until recently observers failed to detect any carbon-dioxide production in nerve, and this fact made many lean to the view that nerve was an inert conductor. More recently Tashiro (113), Parker (96, 97, 98), and Fenn (30) have shown that minute quantities of carbon dioxide are produced by nerve even at rest, and that this quantity is measurably increased during activity, as is also the oxygen consumption (31).

The heat production of muscle is well known and has been extensively studied with quantitative methods by A. V. Hill and his associates. Recently, with refined methods Downing, Gerard, and Hill (28) demonstrated heat production in nerve. From experiments with various frequencies of stimulation, Gerard, Hill, and Zotterman (54) inferred that the amount of heat from a single impulse traversing a gram of nerve is a millionth of a calorie. This is very much less than the heat production in muscle, which is of the order of 0.003 calorie per gram for a single functional response. In the matter of heat production there is a further qualitative similarity and quantitative difference between nerve and muscle, in that each tissue shows an immediate heat production, which results during functional activity, and a delayed heat production, which occurs at a slower rate but lasts for several minutes after activity has ceased. In the case of muscle contracting for a short time, the delayed heat production amounts to about half the total. In the case of nerve the delayed heat production amounts to about nine-tenths of the total and lasts as long as ten minutes after the cessation of activity (28). Clearly the activity of nerve, as well as of muscle, is metabolic.

FATIGUE

One of the most striking apparent differences between nerve and muscle is in relation to fatigue. If an isolated muscle be made to contract by a rapid series of stimuli, it will soon become obviously fatigued and the ability to contract will fall oft to a small fraction of its initial value. A nerve, on the other hand, will conduct a hundred impulses or more a second and continue to do so for hours without apparent exhaustion. From this fact it has been inferred that nerve was unfatigable. But recently Field and Brücke (32) showed that with continued stimulation the refractory phase of a nerve is greatly prolonged. Gerard (50) showed that oxygen consumption and heat production in nerve become less if stimulation is continued at high frequency than if it is intermittently interrupted with periods of rest. In some recent experiments (Forbes and Rice, 44*a*), the action current of nerve has been found to grow smaller in a prolonged series of responses at high frequency (73). Gerard (51) emphasized the fact that, in the case of nerve, continued activity at a given frequency of stimulation does not result in a decline of activity to the point of exhaustion, but that an equilibrium is reached at a reduced level of functional activity, and this level may be maintained for a long time. Increase in the frequency of stimulation causes the functional level to fall farther and attain a new equilibrium; decrease in frequency causes a partial restoration to a higher equilibrium level. Gerard proposes the term *"equilibration"* fatigue to signify the difference between this condition and the type of fatigue that progresses to exhaustion. The fact that fatigue can be demonstrated in nerve tends to make the difference between nerve and muscle quantitative rather than qualitative in this respect, as well as in other features of functional activity.

Gerard (52) has attempted to bring together the facts which have emerged from the chemical and dynamic studies of nerve in a provisional theory. He shows how a cycle of three related reactions might give rise to the known phenomena of refractory phase and equilibration fatigue.

We have considered in some detail the functional properties of nerve and muscle. Of the excitable tissues, medullated nerve and striated muscle are the two most highly differentiated for the special functions of conduction and contraction respectively. It has been shown that, in spite of their differences, the essential nature of the conduction process is apparently the same in both. That, in tissues so different in structure and embryonic ancestry, these essential elements of function should be the same, is a matter of profound significance. It suggests the inherent propensity to react in this way as perhaps a basic property of living cells in general.

An important conclusion which follows from the all-or-none character of the nerve-impulse, and which is not yet as well recognized as it might be, is the fact that, whatever the functional properties of the gray matter may be, the nervous system must utilize impulses of the kind which we have been describing wherever they are transmitted through axons, whether in the peripheral nerves or in the white matter of the brain and cord (87). It has been customary in the past to treat nerves as if they were pipe lines in which continuous streams of nervous energy could be graded as to intensity, much as one grades the flow of water by opening and shutting a faucet. It is evident that this mode of treatment is fundamentally wrong. We must recognize that the impulses mediating sensation in the afferent nerves, the impulses in the axons of the internuncial neurons of the entire central nervous system, and finally the impulses which traverse the efferent nerves to evoke responses in muscle or gland, are all successive disturbances of the transient, all-or-none type.

Muscular Contraction

It has already been noted that nerve and muscle are differentiated and specialized for the two distinct functions, conduction and contraction, and yet that fundamental properties of the propagated disturbance are the same in both. It remains to mention the salient properties in which muscle differs from nerve. The conduction process in muscle is incidental to its prime function, contraction. Hill says, "a muscle is essentially a mechanism for transforming chemical into mechanical energy" (60).

Anatomically, skeletal muscle is made up of fibers varying roughly from 20 to 80 microns in diameter, each consisting of a large number of *myofibrils.* The myofibrils show a banded or striated structure, due apparently to alternate transverse layers of two kinds of material called isotropic and anisotropic.

It is noteworthy that whereas the propagated disturbance set up by a single stimulus, as measured by the action current, attains its maximum in less than 2σ (at mammalian body temperature) and is practically over in 10σ, contractile tension requires about 10σ to reach its maximum and lasts altogether for 20σ or more. Sherrington (107) perfected a torsion-wire myograph for recording accurately and without lag the tension developed in isometric contraction, i. e., contraction when the muscle is not allowed to shorten and the contractile effort is revealed as tension. His observations brought out the striking fact that in the simple twitch (response to single stimulus), after the tension has reached its maximum and remained constant for a brief time, it *suddenly* begins to decline. The moment when the decline begins is revealed by an "angle" in the myograph record. Typical records from the same muscle at two temperatures are shown in Figure 7, together with the action currents marking the underlying propagated disturbance. The method of analysis made possible by this refined isometric recording has been most fruitful in the study of reflexes. Fulton (46) has used it with great ingenuity in determining the degree of synchronism in onset and cessation of contractile activity in the fiber groups of the various muscles involved in spinal reflexes.

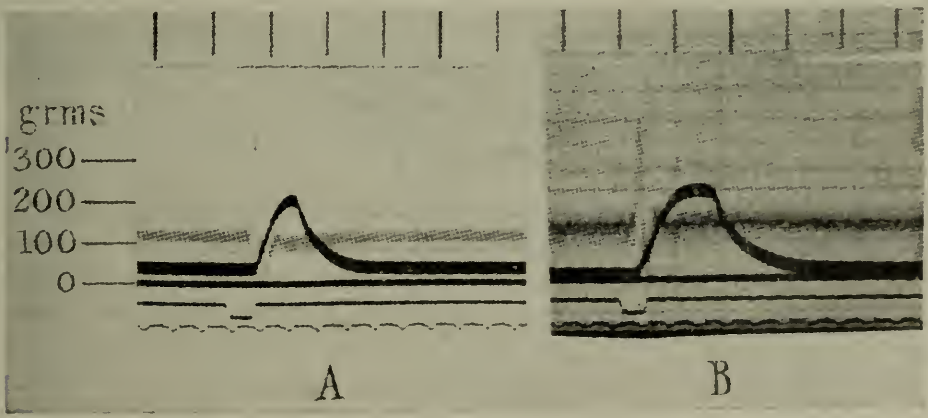


FIGURE 7

ISOMETRIC RECORDS OF THE TWITCH OF A GASTROCNEMIUS MUSCLE OF FROG TOGETHER WITH STRING GALVANOMETER RECORDS OF THE ACTION CURRENT

The upper curve shows the action current; the second curve (darker) records tension developed by the muscle. The tension scale is shown at the left. Below is a signal showing the brief opening of a short-circuiting key, whereby the stimulus from a vibrating interrupter was permitted to reach the nerve. Time shown above in 0.04-second intervals. *A,* temperature 24.5°C.; initial tension 35 gms. *B,* temperature 16°C.; initial tension 15 gms. (From: Fulton, J. F. Some observations upon the electrical responses and shape of the isometric twitch of skeletal muscle (intact). *Proc. Roy. Soc. (B),* 1925, **97**, facing 430.)

It is an old observation that a muscle will contract more powerfully if subjected to a moderate initial load. "The maximum work can be obtained only by variable load, which at every length during the contraction is exactly equal to the maximum tension the muscle can exert at that length" (58). The extensive researches of Hill and his colleagues (57, 58, 59, 60, 61, 63) have solved many problems in the thermodynamics of muscular contraction. Heat production occurs in several distinct stages. "Immediately on stimulation there is a large and rapid evolution of heat; this, in a contraction of short duration, represents about half the 'initial' energy: it is complete long before the moment when the maximum tension is attained; then follows relaxation, accompanied by the evolution of the other half of the 'initial' heat. In a prolonged contraction there is another equally obvious phase, that associated with a constant rate of heat-production, lasting as long as the contraction is maintained" (60). As already mentioned, this initial heat production is followed by a recovery heat production, very much more gradual but slightly larger in amount, and lasting in muscle as long as 15 minutes. The initial heat production occurs, as does contraction, even if the muscle is isolated in an atmosphere deprived of oxygen. The recovery heat production occurs only in the presence of oxygen. Hill concludes that the initial energy liberation is non-oxidative, but that recovery is an oxidative process. He has likened the recovery process to a combustion engine driving a dynamo which recharges a storage battery.

Chemically it has been established that glycogen, the fuel which supplies the energy of contraction, is converted into lactic acid and sodium lactate. Part of the lactic acid is oxidized and the resulting energy is utilized in converting the remainder into glycogen. The portion oxidized is estimated to be one-fourth of the total. Carbon dioxide is ultimately given off as a result of the oxidation. Other reactions probably occur, but the above indicates in outline those most definitely established.

Central Nervous Functions

Let us now turn to a consideration of the special functions of the central nervous system. The gray matter of the brain and cord comprises the adjustor mechanism. Its functional properties can best be approached through a study of the simplest reflexes involving the synapse between neurons.

The contrast between nerve-trunk and reflex arc may be viewed from the angle of the diversity of purpose which the structures serve. The nerve-fiber apparently exists for the purpose of transmitting messages to remote parts, rapidly, economically, and without modification. The central structure appears to serve as a junctional point where messages from many regions may be correlated, relayed, and distributed to other regions. In this respect the fibers and centers may be likened to the wires and central offices, respectively, of a telephone system.

Anatomically the nerve-fiber is characterized by its comparative simplicity, extending as it does for great distances without branching or changing much in size or form. In contrast with this the gray matter presents a picture of the most prodigious complexity, with intricate end-branches and dendrites providing the connections for extensive coordination of conducting paths.

Finally, the contrast between nerve-trunk and reflex arc may be viewed from the angle of functional differences. Sherrington in 1906 (102) summarized the differences between conduction in nerve-trunk and reflex arc, the most important of which are as follows. Reflex conduction shows:

a) *slower speed* as judged by latency of response;

b) *after-discharge,* i. e., persistence of response after stimulation has ceased, often for several seconds;

c) *summation,* single stimuli in many reflexes failing to produce any response, whereas a repeated series is effective;

d) *irreversibility,* conduction from afferent to efferent neurons being possible, but in the reverse direction through the central structure, impossible;

e) *fatigue* on continued stimulation, in contrast with the nerve-trunk, which exhibits extraordinary resistance to fatigue;

f) greater *variability of threshold* or ease with which responses can be evoked;

g) far greater *dependence on blood supply and oxygen,* and correspondingly greater *susceptibility to anaesthetics;*

h) *reinforcement* by allied reflex arcs;

i) *inhibition,* afferent nerves having the property of suppressing as well as initiating central activity.

Histological research has not yet settled conclusively whether there is indeed a true surface of separation between connecting neurons, or whether there is actual protoplasmic continuity. The indisputable structural difference between the reflex arc and the nerve-trunk lies in the intricately branching system of connections, on the one hand, and the isolated, unbranched arrangement on the other (79).

Lucas, who did more, perhaps, than anyone to elucidate the fundamental nature of conduction in nerve and muscle, raised the following question (83): "Are we to suppose that the central nervous system uses some process different from that which is the basis of conduction in peripheral nerves, or is it more probable that the apparent differences rest only on our ignorance of the elementary facts of the conduction process? If we had a fuller knowledge of conduction as it occurs in peripheral nerve, should we not see Inhibition, Summation, and After-discharge as the natural and inevitable consequences of that one conduction process working under conditions of varying complexity?"

The scope of this idea is large. A reduction of the elements of neural activity underlying consciousness and behavior to the single basis of the nerve-impulse, many of whose physical properties are now known, would be a generalization comparable to the reduction of all the various chemical elements to their constituent protons and electrons.

Lucas further suggested: "that we should inquire first with all care whether the elementary phenomena of conduction, as they are to be seen in the simple motor nerve and muscle, can give a satisfactory basis for the understanding of central phenomena; if they cannot, and in that case only, we shall be forced to postulate some new process peculiar to the central nervous system."

A task of cardinal importance in the study of the nervous system is to determine the extent to which the peripheral type of conduction can serve to explain the known facts of central or reflex function. Does every neural disturbance sweep over the conducting paths open to it, leaving them refractory, so that sustained activity must be intermittent? Or is there something at the synapse which can pass into a sustained state of activity uninterrupted by refractory phase and capable perhaps of gradation?

Spinal Reflexes. In mammals the reflexes most accessible to study are the spinal reflexes, involving contraction of the limb muscles. It is possible to isolate the local centers which coordinate the movements of the hind limbs by transecting the spinal cord in the thoracic region. In that way the lumbar enlargement of the spinal cord, which contains all the motor neurons innervating the muscles of the hind limbs, is cut off from the influence of the brain and higher centers in the cord. This operation makes it possible to study the extent to which a local coordinating mechanism exists in these lower centers. Transection of the spinal cord between the brain and the cervical enlargement containing the local centers for the fore limbs makes possible the study of the local reflexes in both the hind limbs and the fore limbs and the coordination of the two pairs of limbs with each other. This operation destroys the respiratory center and makes it necessary to employ artificial respiration to keep the animal alive. Transection of the brain-stem, just in front of the cerebellum, eliminates the entire cerebrum and with it consciousness and spontaneous activity. It does not interfere with spontaneous respiration or with certain coordinating effects of the lower centers of the brain, and it permits the study of spinal reflexes under conditions which are in some respects more nearly normal than those obtaining when the spinal cord is severed below the medulla oblongata. This operation is known as decerebration. The most striking immediate effect is the development of extensor rigidity, a condition described and studied by Sherrington. All the extensor muscles of the limbs go into a state of sustained contraction and remain in this condition for many hours, in fact as long as the animal can be maintained in a normal condition as regards temperature and circulation. As Sherrington (102, p. 302) has pointed out, the extensor muscles are those which support the animal's weight against the force of gravity and their sustained contraction is to be looked on as a postural reflex. Its reflex nature has been demonstrated by the fact that any interruption of the path from the receptors of muscle sense, located within the muscles and tendons, to a nerve-center in the hind-brain and back to the extensor muscles through their motor neurons, abolishes the rigidity.

All three of these operations have been used extensively by Sherrington (102) in the study of spinal limb reflexes. The initial operation is performed under deep surgical anaesthesia and thereafter the reflexes may be studied after the local centers have been freed from the effects of the anaesthetic. In the case of decerebration or transection of the cord at its junction with the brain, the cerebrum, which is the center of consciousness, is destroyed, and the animal thereafter becomes a reflex automaton in which the reflex mechanism may be studied without the possibility of pain. In the case of low spinal transection in the thoracic region, there is no nervous connection between the lumbar centers coordinating the muscles of the hind limbs and the rest of the central nervous system; thus the hind limb reflexes may be studied without provoking any sensation whatever in the animal.

These methods have yielded a great mass of information about the mammalian limb reflexes (104). Only the more constant and important of these reflexes will be described here. When a skin surface or a sensory nerve in the hind limb is strongly stimulated in either a spinal or decerebrate animal, the regular reflex response is a rapid flexion of the hip, knee, and ankle in the stimulated leg. Usually this is accompanied by a slightly more gradual but equally strong extension of the same three joints in the opposite hind leg. These two reflex responses are called the *flexion reflex* and the *crossed extension reflex*. Their purposive significance has been pointed out by Sherrington (102). If a normal animal steps on a thorn or otherwise brings his foot in contact with an injurious object, the natural defensive reaction is withdrawal of the foot, which is best effected by flexion of hip, knee, and ankle. That the body may be supported during this reaction requires the contraction of the extensor muscles of the opposite hind limb. The crossed coordination whereby flexion in one hind limb is accompanied by extension in the other is found not only in the defensive reaction just described but also in normal progression. A walking animal such as a cat or dog normally flexes the joints of one leg while extending those of its mate; rhythmical progression movements, in which this coordination is beautifuly demonstrated, can easily be evoked in a cat or dog after low spinal transection. Many of Sherrington's observations have shown a functional linkage between the flexor motor center of each hind limb and the extensor motor center of its mate.

One of the most interesting and striking features of these limb reflexes is the reflex inhibition shown in the reciprocal innervation of antagonistic muscles. If the flexor and extensor muscles of the knee joint contract simultaneously, they work against each other. Sherrington (102, 104, 105, 106) has shown that reflex excitation of one group of muscles is accompanied by reflex inhibition of the antagonistic group, which insures their relaxation and the elimination of wasteful opposition. Thus in the flexion reflex there is not only reflex excitation of the flexor muscles but simultaneously a reflex inhibition of the extensors. This is especially well demonstrated during decerebrate rigidity, in which, as has been already stated, the extensor muscles are in a state of sustained contraction. If an extensor muscle be isolated it is found that when the powerful afferent stimulus which evokes the flexion reflex is applied, the extensor muscle relaxes instantly; this is due to central inhibition.

Other examples of inhibitory action are known elsewhere in the body, as for example the inhibition of the heart by the vagus nerve. The stimulation of the vagus causes a slowing or cessation of the heart-beat; this occurs even when the vagus nerve is severed from the central nervous system and the peripheral portion stimulated. There is an important difference between such peripheral inhibition and reflex inhibition of skeletal muscles. The heart muscle has the property of spontaneous contraction; therefore peripheral inhibition is a necessity if the heart-beat is to be slowed to less than its spontaneous rate. Striated muscle has no tendency to spontaneous contraction; normally it contracts only when excited by the discharge of impulses from the central nervous system. If motor nerve-impulses cease, muscular contraction ceases. There are no special inhibitory impulses set up in the motor nerves to stop contraction in the skeletal muscles (115, 45); reflex inhibition consists in stopping the discharge of motor impulses from the reflex center (22).

There are certain interesting differences between the reflex contractions of flexor and extensor muscles (74). The flexion reflex usually starts very quickly, and if the stimulus be strong all the motor neurons which participate in the reflex begin discharging impulses at once; thus the muscular contraction has a brisk and rapid onset not differing much from the twitch evoked by applying a maximal stimulus to the motor nerve. Forbes and Gregg (40), stimulating an afferent nerve with single induction shocks and recording the action current in the motor nerve which innervates the flexor muscle in the hind leg of a cat, showed that the time consumed by the disturbance traversing the spinal cord is only about 4σ. Forbes and Miller (42), recording the action current which marked the entrance of afferent impulses into the brain, found that when a single stimulus was applied to the sciatic nerve of a decerebrate cat, the first impulses reached the medulla oblongata about 8σ after the stimulus was applied, or not less than 6σ after the afferent impulses had reached the spinal cord. From these observations it is evident that in the flexion reflex the motor response is actually initiated by the spinal cord an appreciable time before the afferent impulses make the animal aware of the stimulus by reaching the cerebral cortex (39). Such prompt and uniform conduction though the arc of the flexion reflex forms the basis of a machine-like regularity of response, apparently not differing very greatly from that found in the nerve-muscle preparation, in which the motor nerve is stimulated and the muscle responds without the interposition of any reflex mechanism.

In the case of the crossed extension reflex the condition is quite different. This probably involves more internuncial neurons than the flexion reflex; at all events, the conducting path has to cross the median plane of the spinal cord. The reflex is more variable than the flexion reflex and is much less regularly evoked by single stimuli applied to the afferent nerve. It is frequently impossible in decerebrate animals to evoke this reflex except by a series of several successive stimuli. Its latency is considerably longer, the time occupied in traversing the gray matter (*reduced reflex time*) being never less than 12σ (38). It never begins with a full discharge of the center such as is found in the flexion reflex, but instead its onset is gradual. Liddell and Sherrington (74) have shown that the first few stimuli only evoke response in a relatively small percentage of the motor neurons which ultimately take part if the stimulation is continued long enough. Gradually more and more motor neurons are brought into action as the reflex develops. They call this process *recruitment* (75). The after-discharge of the crossed extension reflex is much more prolonged than that of the flexion reflex. In the decerebrate animal this is partly due to the secondary reflex effect caused by stimulation of the intramuscular receptors (muscle sense) which set up a *proprioceptive reflex* (102), resulting in a sustained contraction of the extensor muscle. Even when the afferent part of the proprioceptive reflex arc is interrupted by severing the dorsal roots of the spinal nerves, the crossed extension reflex shows a more prolonged after-discharge than does the flexion reflex. The crossed extension reflex usually disappears first in a moribund animal. It appears to involve a more delicate and vulnerable mechanism than that which mediates the flexion reflex.

Let us now consider the specific differences between nerve-trunk and reflex conduction appearing in the facts now known and, following Lucas' suggestion, see whether any of them create insuperable obstacles in the way of explaining reflex phenomena in terms of the all-or-none type of conduction established for peripheral nerve. It is granted that the reflex utilizes the all-or-none type of response in the axon; the question is whether a similar type of response accounts for conduction through the synapse.

Conduction Time. Jolly (67) found the reduced reflex time of the knee-jerk to be about 2σ, i. e., about half as long as that of the flexion reflex. The actual distance traversed within the cord is very short; the velocity of conduction in peripheral nerve is about 100 meters a second, or 10 centimeters per sigma; therefore it is evident that conduction through the central part of the reflex arc is much slower than in peripheral nerve. This fact, however, does not demand a qualitatively different kind of conduction; the difference may be purely quantitative. Rapid conduction is a highly specialized property of medullated nerve-fibers. It is altogether likely that the finer terminal branches and dendrites in the synaptic region resemble more closely the primitive tissues as regards their speed of conduction, and yet there may be no essential difference in the conduction process. The very much greater delay in the reflex response to weak stimuli, described by Sherrington, amounting, in the case of the scratch reflex, to two or three seconds, involves a different principle, for in these cases the response was evoked only by repeated stimulation. The problem thus becomes one of the summation of propagated disturbances and will be dealt with as such presently.

After-Discharge. The continued discharge of impulses in the motor neurons for several seconds after the cessation of afferent stimulation, especially notable in the crossed extension reflex, appears at first sight to be at variance with the principle of peripheral conduction in which the disturbance sweeps over the tissue, leaving it refractory. There appears to be instead a condition of sustained activity. But when we consider that in the gray matter we are not dealing with isolated, unbranched paths, as in the nerve-trunk, but with a complex system in which one afferent fiber is probably connected with many central neurons through extensive branching, we see that this sustained activity need not involve any fundamentally different conception of function from that developed in the case of the peripheral tissues. We need assume only sufficiently elaborate and extensive paths including chains of neurons, each adding a considerable measure of synaptic delay, to account for any observed after-discharge without introducing a functional capacity in any way at variance with the all-or-none principle. The most prolonged after-discharge of the crossed extension reflex would require a sort of central reverberation, that is, some central paths would be traversed repeatedly, but this does not necessarily require the assumption of a qualitatively different function.

Summation. Many reflexes cannot be evoked by a single volley of afferent impulses; a succession is required—sometimes as many as forty or fifty stimuli must be applied to bring about the reflex contraction (102). In the normal nerve-muscle preparation a single adequate stimulus evokes a full-sized response. Adrian and Lucas (12) were able to find conditions which made possible a summation of propagated disturbances in the nerve-muscle preparation. If the neuromuscular junction was sufficiently fatigued, a single impulse in the nerve failed to evoke contraction in the muscle. But if a second stimulus was applied during the supernormal phase of recovery following the relative refractory period, a contraction could be evoked. They were also able to duplicate this effect by interposing a narcotized region in the nerve between the stimulating electrodes and the muscle. Their original interpretation in terms of decremental conduction must be revised; and as long as we have no exact knowledge of the physical nature of the nerve-impulse we cannot hope to form a precise picture of the mechanism of summation. The most likely explanation of the observed facts is that in the supernormal phase of recovery there is, correlated with increased excitability, an increased size of response. The acid condition which Adrian found essential to a supernormal phase is what one would expect in a fatigued nerve-muscle preparation, in which there is known to be accumulation of lactic acid. Adrian and Lucas (12) suggested that the type of summation they found in the fatigued nerve-muscle preparation might serve as a clue to summation in the reflex arc. Adrian (3) further drew attention to the fatigability of the reflex arc and its dependence on oxygen as indications that acidity may be a common, if not a constant, condition in the neighborhood of the synapse. These considerations show that the facts of reflex summation do not necessarily force us to postulate a different kind of function in the synapse from that found in the nerve-trunk.

Irreversibility. The synapse has a valve-like property which enables impulses to pass from afferent to internuncial or motor neurons but not in the reverse direction. Lillie (76) offered an explanation for this, based on the assumption that the action current is the essential cause of propagation of the nerve-impulse. He suggests that irreversibility depends on the relation between the duration of the action current at one point in the conducting path and the chronaxie of the next point, which must be excited by that action current in order that the disturbance may go on. We may assume, for example, that the time relations of excitation and response in the dendrite, which conducts impulses from the synapse to the axon of the next neuron, are much briefer than those of the end-branches through which the impulses approach the synapse. In order that conduction should be irreversible, it is necessary only that the action current in the end-branch should last long enough to excite the dendrite, but that the action current of the dendrite should be too brief to excite the branches of the afferent neuron, because of their longer chronaxie. This suggestion, though speculative, serves to show how irreversibility might be achieved without the necessity of any functional property other than protoplasmic conduction as shown by the nerve-trunk.

Fatigue. It is easy to demonstrate fatigue in the reflex arc. Its fatigability contrasts strikingly with the relative unfatigability of the nerve-trunk. Gerard and Forbes (53) recently showed that fatigue in the flexion reflex was apparently of the equilibration type, the size of response falling to a reduced level depending on the frequency with which the responses are evoked; thus the reflex arc qualitatively resembles the nerve-trunk. Quantitatively there is a great difference, the degree of fatigue being very much greater in the reflex arc.

Sherrington (102, p. 218), Lee and Everingham (71), and Forbes (33) gave evidence of various sorts tending to show that the fatigue in the reflex arc occurs in the particular channel of approach to the motor center, rather than in the discharging motor neurons themselves. This tends to place fatigue in the synapse, which is the structure to which Sherrington ascribed most of the functional properties peculiar to the reflex arc (102, p. 16). Stiles (111) has drawn attention to the fact that the finer branches of the neurons in the synaptic region are highly attenuated and that therefore we should expect any material required for the transmission of the impulse to be more rapidly exhausted here than in the larger fibers.

Variability of Threshold. Sherrington emphasized the fact that the threshold of reflex excitation is more variable than that of the isolated nerve. This appears paradoxical. In view of the all-or-none law of the nerve-impulse, it is hard to see how the threshold of the reflex can be other than the threshold of the afferent nerve-fibers. Lutz (84) compared the threshold for reflex contraction on stimulating the afferent nerve in the frog with the threshold for muscular contraction when the stimulus was applied directly to the motor nerve in the same animal. When the animal was cooled, the threshold for the reflex rose nine times as much per degree centigrade as did the threshold of the nerve-muscle preparation. The reflex threshold presumably depends on the number of afferent fibers excited and is not the same as the threshold of the most excitable fiber among them (35, p. 377). This consideration emphasizes the importance of the branched character of the central connections. The treatment of a reflex arc as consisting typically of a single afferent fiber connected through a synapse with a single motor fiber is misleading in that it ignores the extensive interconnection of many afferent neurons with many motor neurons. The supposition that a large number of converging impulses give a more effective stimulus to the motor neurons than a smaller number, although suggesting a type of central activity of graded intensity instead of one of all-or-none character, does not necessarily force us to assume a new type of process. A portion of the synaptic region common to all the afferent paths might have a very brief refractory phase. This would account for the effect of convergence, on the basis of summation of successive disturbances, for even the synchronous volley of impulses set up in the afferent nerve by a single shock would probably arrive at the common path at slightly different times.

Dependence on Blood Supply and Susceptibility to Anaesthetics. Reflex centers are strikingly dependent on blood supply as compared with peripheral nerves. A concentration of ether or other anaesthetic in the blood, sufficient to abolish all ordinary reflexes, does not appreciably affect the function of the peripheral nerves. On the other hand, a nerve without its blood supply will ultimately lose its function, and anaesthetics in sufficient concentration will abolish function in a nerve-trunk. Thus it appears that, both in dependence on blood supply and oxygen and in susceptibility to anaesthetics, the difference between the reflex arc and the nerve-trunk is only one of degree. It has been shown (35, p. 386) that there is an adequate basis for the observed differences in the known properties of the synaptic part of the conducting path.

Mutual Reinforcement between Allied Arcs. Sherrington (102) has given several instances of reinforcement of the activity of one reflex arc by stimulation of an allied arc. This phenomenon appears to fall readily into the category of effects just discussed under variation of reflex threshold. If those effects can be explained by convergence of individual afferent paths, so can reinforcement. Camis (19) found that contraction in the flexion reflex was greater, if evoked by simultaneous stimulation of two nerves containing afferent fibers from different parts of the hind foot, than could be evoked by stimulating either nerve alone. This might be ascribed to the connection of the afferent fibers of the two nerves with different motor neurons, but there is evidence from other sources (24) that both sets of afferent paths are connected centrally with many of the same motor neurons. Central reinforcement may depend in part on additional motor neurons involved in the reinforcing reflex, but it probably also depends in part on convergence of conducting paths from various sources at common points in the nervous system.

Reflex Inhibition. This phenomenon apparently involves something wholly different from ordinary conduction, for, whereas the nerve-impulse ordinarily induces functional activity, the central inhibitory process suppresses activity. The afferent impulses entering the spinal cord cause a cessation of the discharge of motor impulses from the center.

In 1885 Wedensky (116) described a phenomenon in the nerve-muscle preparation, bearing at least a superficial resemblance to reflex inhibition, and in 1911 Lucas (82), in a brilliant series of experiments, established the true nature of the Wedensky effect. Lucas further indicated a possible way in which the principle involved might conceivably apply to reflex inhibition.

The Wedensky effect is as follows. If a nerve-muscle preparation is moderately fatigued by stimulation through the motor nerve, a stage is reached in which a series of stimuli of proper strength and frequency applied to the nerve evokes only an initial twitch in the muscle, followed by its complete relaxation. While these stimuli are being applied, additional stimulation of the nerve nearer the muscle fails to make it contract. If the frequency of the stimuli is decreased to a certain point, tetanic contraction appears. It will be recalled that during the recovery of nerve from a previous impulse there is a relative refractory phase, during which a second response will be of subnormal size. Lucas showed that the Wedensky "inhibition" occurs only when the frequency of stimulation is such that each impulse after the first occurs during the relative refractory phase following the preceding impulse, and therefore is subnormal. The fatigued neuromuscular junction constitutes a partial block to conduction. A full-sized nerve-impulse will excite the muscle, but a subnormal impulse will fail. There is a critical frequency below which the nerve is sufficiently recovered after each response to render the impulses large enough to excite the muscle and thus cause a sustained contraction. Above this critical frequency all impulses after the first are subnormal, the muscle therefore responds with a single twitch to the first impulse and thereafter relaxes. Although Lucas, in discussing this effect, assumed decremental conduction at the junction as the basis for the extinction of the subnormal impulses, it has been shown that this assumption is quite unnecessary (44, p. 605). That the effect depends on subnormal impulses is a demonstrated fact; to explain the observations without invoking decrement requires only the assumption that the threshold of the muscle-fiber at the junction is raised to such a point that a subnormal impulse can no longer excite it, or that fatigue has reduced the size of the impulses in the terminal branches till they are below the muscle's threshold of excitation.

Lucas (82, 83) showed how the same principle might be applied to reflex inhibition (cf. 6). If some part of the reflex path, e.g., the last internuncial neuron leading to the motor neuron, is made to conduct impulses with a frequency which renders them subnormal, and the synapse leading to the motor neuron presents a partial block to conduction, not only will the motor neuron remain unexcited, but the internuncial neuron so occupied will become a block through which no full-sized impulse can be made to pass. In such a neuron there would be, correlated with the condition of the synapse, a critical frequency of nerve-impulses above which the effect would be inhibitory, and below which, excitatory.

It should be noted that, though Lucas first suggested an internuncial neuron and a region of decrement in the synapse to which it led, as a possible mechanism, neither of these is essential in the application of this principle to reflex inhibition (35, 44). The part of the path which is occupied with conducting impulses of high frequency might be a terminal branch in the gray matter, and since certain reflex paths are regularly inhibitory in action, we might assume that it is a regular property of such a branch to respond at a characteristic rhythm to excitation of whatever sort. Transformation of rhythm such as this implies has its counterpart apparently in the autonomic ganglia (100). Decremental conduction was generally assumed when the earlier discussions of this subject were written, but has since been disproved. On the other hand, Davis, Forbes, Brunswick, and Hopkins (26) showed on theoretical grounds that a transitional decrement must exist and almost certainly extends over distances as short as those involved in synaptic junctions. Furthermore the idea of decrement is not essential to the schema, since a high threshold of excitation at the synapse, rendering the motor neuron inaccessible to subnormal impulses, would suffice just as well. Querido (100) has reported a phenomenon of smooth muscle innervation, which apparently exemplifies the Wedensky effect as a normal relation between impulse frequency and muscle response.

It should be emphasized that no attempt is being made to decide on the *probable* central mechanism. The above hypotheses are set up merely to test the question raised by Lucas as to whether any insuperable obstacles stand in the way of interpreting reflexes on the same ultimate basis as conduction in the nerve-trunk.

Other Reflex Effects. A number of important facts concerning spinal reflexes must be accounted for by any hypothesis relating to the mechanism in the spinal centers. Sherrington (103) observed a balancing of the antagonistic effects of excitation and inhibition by stimulating simultaneously afferent nerves from opposite sides of the animal with graded strengths of stimuli. The balancing of opposite central effects seems to show what he called an *algebraic summation,* the degree of muscular contraction being intermediate between the full contraction evoked by the excitatory stimulus alone, and the relaxation caused by the inhibitory stimulus alone. This effect has been especially studied in the case of the extensor muscle, which is excited in the crossed extension and inhibited in the flexion reflex. It appears that a maximal stimulus, that is, one strong enough to excite all the fibers, applied to an afferent nerve evoking the flexion reflex, always produced complete inhibition of the extensor, and no strength of excitatory stimulus could break through it. Algebraic summation apparently occurs only when some of the afferent fibers are not excited; it apparently depends on leaving some of the extensor motor neurons uninhibited (34).

Various reversals of reflex effect are described by Sherrington and his co-workers (109, 110). Interesting examples are the change from reflex excitation to reflex inhibition under the influence of moderate chloroform anaesthesia, and the postural reversal which Sherrington (102) and Magnus (86) have found in certain reflexes. The type of response to a given stimulus is in some cases determined by the posture existing when the stimulus is applied. Thus Sherrington has been able under certain conditions to evoke flexion in an extended limb by the same stimulus which evokes extension in a flexed limb. These effects have been discussed (35, p. 396) in connection with Lucas' question as to whether reflex phenomena conform to the principles of conduction as manifested in peripheral nerve, and no clear evidence was found in them demanding the assumption of a qualitatively different function in the reflex arc.

SHERRINGTON'S HYPOTHESIS

Since the publication of the above-mentioned discussions, in which the attempt was made to see whether the functions of the reflex center could be interpreted in terms of the nerve-trunk type of conduction, Sherrington (108) has proposed an alternative interpretation. He emphasizes the conspicuous property of summation in the reflex arc, whereby repetition of the afferent impulse renders a subliminal effect supraliminal. He says, "This result is explicable if at some central situation there be a structure which is something other than a nerve-fiber, and has, unlike nerve-fiber, no absolute refractory phase, and has, as have many other cell-structures including skeletal muscle-fiber itself, a property of summating its successive reactions when these are not too far apart in time." He further points out that the ionic change, which on the membrane theory underlies the progress of the nerve-impulse, may be expected on reaching the limiting membrane, which may well exist at the synapse, to lead to a different condition from that which obtains in the axon. This suggestion implies that successive impulses, on reaching the synapse, cause a cumulative ionic concentration, which is impossible in the continuous nerve-fiber. Sherrington denotes the "exciting state or agent" *E,* but is noncommittal as to its nature. He assumes merely that it differs from the nerve-impulse in that it may be graded in intensity, being built up by successive impulses, outlasting them and subsiding more slowly.

To explain inhibition, Sherrington postulates "a state or agent which produces lessening or total disappearance of *E*," and this he designates *I.* He suggests that this might be of the nature of diffusion or adsorption of the excitation-ions, or "a neutralization of them, chemical or electrical." Reflex inhibition, on this view, would depend on the fact that impulses in the inhibitory afferent nerve, on arrival at the synapse, set up the state *I,* which could be graded in amount according to the number of afferent impulses, just as *E* could be graded in the excitatory reflex arc. This hypothesis can be very simply applied to the observed facts of balancing ("algebraic summation") of the opposed reflex effects, excitatory and inhibitory, in the spinal center, as well as to the facts of summation and after-discharge.

At present there seems to be no crucial evidence to settle the question whether synaptic function depends on the arrangement and distribution in time and space of impulses qualitatively like those in peripheral nerve, or whether it involves essentially different functional properties, such as

Sherrington has suggested.[2] In selecting a working hypothesis, simplicity is to be desired. In one sense Lucas' quest for an interpretation of reflexes which invokes no functional properties but those known in the nerve-trunk would, if successful, be a simplification of the highest order. In another sense, Sherrington's proposal is the simpler, for on the alternative view a number of subsidiary hypotheses are needed to bring the behavior of the synapse in line with that of the nerve-trunk. Specifically the fact that a single stimulus, setting up a synchronous volley of impulses in the fibers of the afferent nerve, suffices to produce reflex inhibition of the extensor muscle, forces us, on Lucas' interference theory, to assume a special mechanism to establish the proper rhythm of impulses in the center. This mechanism might be an attenuated part of the central path, having so brief a refractory period that a single impulse approaching through the main axon could induce in it a rapid succession of impulses. Such a structure with a characteristic rhythm of responses, followed by a region of high threshold or "transitional decrement," would provide the mechanism needed for the Wedensky type of inhibition (35, p. 410). On the other hand, the variation in the frequency of discharge which seems characteristic of central nervous structures and the change of the duration of refractory period on repeated excitation are facts which may be cited against the probability of this type of inhibition. It is difficult to picture the constancy of rhythm, of size of impulse, and of threshold requisite to keep an inhibitory arc true to its function.

It is interesting to consider whether peripheral inhibition, which is found in the heart and certain smooth muscles, may be interpreted on the interference theory. Querido's observations, already mentioned, suggest that they may be. Howell (65) has compared the inhibitory processes in various tissues and has concluded that they are fundamentally the same in all.

In favor of Sherrington's proposal may be mentioned the fact that though graded activity and accumulation of effect are foreign to the character of the impulse in medullated nerve, they undoubtedly are common properties in other protoplasmic systems. Summation of successive nerve-impulses is strikingly illustrated in the smooth muscle of vertebrates. Lapicque (70), Querido (100), and many others have furnished examples in which several nerve-impulses were required before the muscle would contract. The local excitatory process in both nerve and muscle appears to be of the nature of ionic concentration which may be graded in amount. The central structure may well be adapted to permit such a graded concentration of ions or whatever other physical condition may underlie excitation. The behavior of sensory receptors suggests that they are structures in which graded and cumulative excitatory conditions are possible. Furthermore it seems probable that the more primitive cells such as the protozoa and perhaps the primitive muscle-cells of the lower metazoa, possess graded functional activity rather than responses of the all-or-none type. It may be that the graded response represents the more fundamental protoplasmic property, and the all-or-none type of response has evolved from it only in the course of the highly specialized differentiation of medullated nerve and striated muscle.

---

[2] Lapicque, in the Dunham Lectures at Harvard, has proposed an interpretation of reflex functions in terms of the relation between the chronaxies of the various neurons. This is distinct from either of the alternative views just considered; therefore our choice may not be limited to these two.

Tonus

The extraordinary persistence of long-sustained contraction in the extensor muscles in decerebrate rigidity has led some observers to infer that it involves a function, "tonus," quite distinct from contraction which causes motion. This problem has been extensively discussed in recent literature (21), (46), (37), and the preponderance of testimony indicates that in the case of skeletal muscle no essentially different function is involved in the sustained contraction from that involved in the execution of movements. There is some evidence that certain autonomic nerve-fibers may innervate skeletal muscle-fibers as well as the so-called somatic motor nerve-fibers which conduct the impulses evoking contraction. This view has been doubted by Hinsey (64). There is further evidence (91) suggesting that if such autonomic fibers do innervate the skeletal muscle-fibers they may possibly exert some physicochemical effect on them, perhaps resembling secretion in gland-cells, an effect which does not directly evoke contraction but which may modify contraction induced through the motor nerves. The evidence of such an effect is of an obscure and uncertain character, and it is premature to accept the conclusion as established.

Gradation of Motor Activity

Since the all-or-none law applies both to the nerve-impulse and to the muscle-response, it is evident that gradation in the strength of contraction must depend either on the number of muscle-fibers in action or the frequency of response in each, or on both (41). Adrian and Bronk (11) have recently determined the frequencies of impulses in the individual motor neuron in the cat in various spinal reflexes. In the flexion reflex the gradation of strength of muscular contraction appears to depend largely on the frequency of impulses; this lies between 5 and 50 per second. In the crossed extension reflex the nerve-impulse frequency is usually in the neighborhood of 70 per second, and gradation depends largely on the number of fibers participating. In the tonic contraction of decerebrate rigidity these authors found the frequency of impulses in the individual fiber to be in the neighborhood of 10 to 20 per second.

It was shown by Forbes and Olmstead (43) that many of the motor nerve-impulses involved in the crossed extension reflex are subnormal, indicating that they fall in the relative refractory period following their predecessors. From this it was inferred that the frequency was over 300 per second. The subsequent observation by Field and Brücke (32), that the refractory period is greatly prolonged on continued stimulation, shows that the inference as to frequency was not justified and reconciles the observations with those of Adrian and Bronk. The muscle-fibers may not always follow the frequency of the motor nerve-impulses, but it seems probable that they generally do.

A typical electromyogram, taken from the muscles of the human forearm during voluntary contraction, is shown in Figure 5*d*. The action currents of the muscle-fiber groups are sometimes synchronized and sometimes not, consequently the resultant composite picture presents an irregular rhythm.

INTEGRATION

The nervous system in higher animals is characterized not merely by the orderly working of definite reflexes, but more especially by the integration of these reflexes into adaptive patterns of behavior. As MacCurdy points out (85), an essential in coordination is speed of conduction.

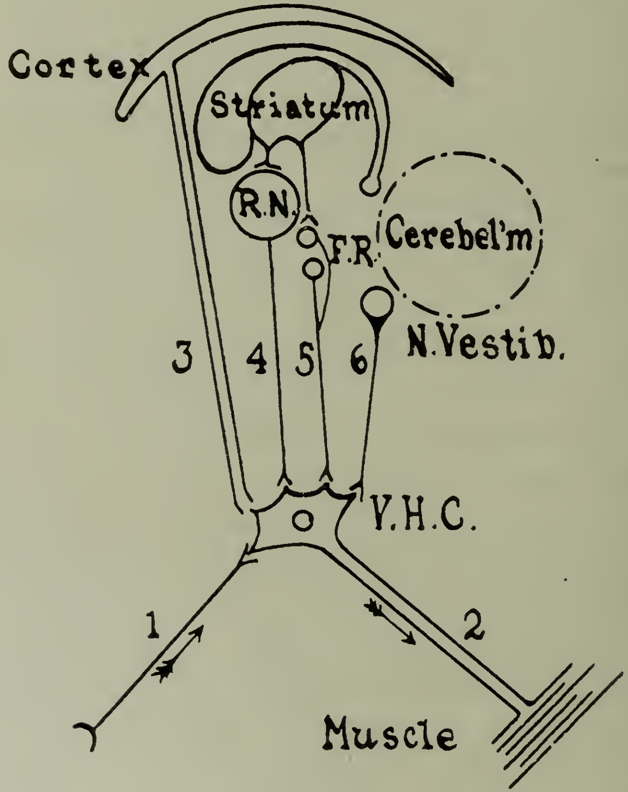


FIGURE 8

DIAGRAM OF THE VARIOUS NEURON LEVELS WHICH PLAY UPON THE MOTOR NEURONS AND CAUSE CONTRACTIONS IN THE MUSCLE

Afferent impulses may come in from a sense-organ and directly affect the ventral horn cell by way of the simple reflex arc (1). The cortico-spinal tract (3) sends down highly integrated impulse-complexes which take precedence over the simpler ones from the nucleus ruber (*R.N.*), the formatio reticularis (*F.R.*), and the nucleus vestibularis (*N. Vestib.*), which traverse, respectively, the rubro-spinal tract (4), the reticulo-spinal tract (5), and the vestibulo-spinal tract (6). But it is the combination of pyramidal (3), extra-pyramidal (4 and 5), vestibular (6), and spinal (1) impulses, all harmoniously playing on the motor cell, that causes normal synergic movement. (From Lennox, W. G., & Cobb, S. Epilepsy. *Med. Monog.*, 1928, **14**.)

That the muscles should efficiently serve the body as a whole, it is necessary that the coordinating impulses should travel from the nerve-centers to the muscles in much less time than the duration of the contractions. This result is achieved by the high speed of conduction in medullated nerve-fibers.

Lennox and Cobb (72) speak of *long-circuiting* and *encephalization* as the evolutionary steps whereby more association neurons, especially those of the brain, are interposed in the reflex path (see Figure 8). These developments provide for delayed response and its modification by virtue of experience, and thus lead to intelligent behavior and abstraction.

Sherrington (102) has described the irradiation of reflexes. As a peripheral stimulus becomes stronger, the central effect becomes greater, which means, in view of the all-or-none law, that more impulses reach the center in a given time. An interesting consequence of this increase is the spread of reflex activity into relatively remote centers in the nervous system. For instance, associated with a powerful flexion reflex is not only extension of the crossed hind limb but extension of the fore limb on the same side and flexion of the fore limb on the opposite side. This coordinated pattern plays a significant part in the act of walking. Other instances of integration are the postural effects described by Magnus (86). Proprioceptive impulses arising in the neck muscles, and depending on their posture, modify the postural reflexes in the limb muscles and to some extent their response to local afferent stimuli. As previously stated, a possible interpretation of postural reversal in terms of Wedensky inhibition has already been proposed (35). MacCurdy (85) has criticized this proposal on the ground that it not only depends on fixed frequencies of impulses in the center but also fails to account for the increased activity of the extensor centers caused by impulses from the hind-brain in decerebrate rigidity. The latter difficulty can be met by assuming that the impulses from the brain approach the motor neurons through another path and thus do not add to the frequency of impulses in that channel of approach wherein the inhibitory interference of impulses is assumed to occur. On the other hand, the effect might be more simply interpreted on Sherrington's assumption of an excitatory state which neutralizes the inhibitory state.

We may conclude that no decisive evidence has been offered to prove that after-discharge, summation, inhibition, and other reflex functions reveal anything which cannot be interpreted in physical terms, either on the basis of Lucas' suggestion as to the arrangement of qualitatively similar nerve-impulses or on Sherrington's suggestion of a balancing of excitatory and inhibitory states.

The introduction of a "vital force" to replace the operation of physical law in the functioning of living matter seems to arise from two causes, the staggering complexity which the behavior of living matter presents and a natural unwillingness to place those faculties which we cherish in ourselves on the same level with machines.

The difficulty arising from the complexity of organisms is due largely to the attempt to make excessively simple explanations fit the behavior of structures whose intricacy should warn us at the outset of the improbability of finding any complete explanation with the means at our disposal. Clearly, with many millions of nerve-cells in the gray matter, each presenting a bewildering array of branching fibers for connection with other cells, there is room for prodigiously complex activity through the mere factor of arrangement, even if the underlying activity is always the same in kind. It is true, as MacCurdy says (85), that "no adequate mechanistic explanation of reflex phenomena has ever been achieved"; but a completely adequate explanation would require virtual omniscience.

The second cause of the "vitalist" point of view rests on a confusion of the means with the end. It seems to me that the only sound method when dealing with the material aspect of living matter, as we must in objective physiology, is to treat it consistently on a physical basis. This position is best expressed by Lloyd Morgan (88), "Monism regards nature and experience as one and indivisible, and all apparent dualism, as a dualism of aspect, distinguishable in thought, but indissoluble in existence." The application of this view to biology has been discussed elsewhere (36), and is admirably set forth in the closing chapter of Hill's *Living Machinery*.

The fact that music is transmitted to the ear by sound waves and nothing but sound waves—physical events not only analyzable but reproducible—does not rob it of its beauty or any other attribute essential to its musical nature. The character of the music depends on the arrangement of the sound waves. To explain how afferent nerve-impulses evoke efferent impulses, we need not assume in the center a vital force which cannot be interpreted physically. But to understand the broader significance of behavior we must recognize the fact of arrangement or integration. Just as the arrangement of sound waves enables them to express something to whose meaning the physical analysis of the waves themselves gives no clue, and as an author uses ink and paper to express thoughts depending on the particular distribution of the ink, so the nerve-impulses and elementary reflexes, by virtue of their integrated arrangement, serve the purposes of conscious life.

SUMMARY

1) The protozoa possess, within the limits of the single cell of which each animal consists, all the functional properties essential to life. In the metazoa, the evolution of the neuromuscular system, which is the chief basis of active behavior, begins with the differentiation of muscle-cells, whose function is contraction. Next appear primitive nerve-cells, which receive stimuli and, being differentiated for the purpose of conduction, transmit impulses to the muscle-cells and thus evoke appropriate motions. From this simple beginning we can trace the development of the nervous system to the vertebrate pattern, in which receptor cells are connected through afferent nerve-fibers with spinal and cerebral nerve-centers (adjustors), and these, in turn, are connected through motor nerve-fibers with muscle-cells and other types of effectors.

2) The effectors, by which the animal reacts upon its environment,

are of various types, the most important of which are ciliated cells, glands, and muscle-cells.

3) The functional properties upon which the working of the neuromuscular system depends are excitability, conduction, and (in the case of muscle) contraction. Excitability and conduction are common to nerve and muscle. In spite of the structural and functional differences between these two highly specialized tissues, the fundamental nature of the functional response is the same in both. This community of function is shown by the nature of electrical excitation, by the electric response, or action current, which marks the progress of the functional disturbance along the active fiber, by the refractory phase, which follows the disturbance, and by the all-or-none character of the response.

4) The all-or-none law is an important feature of the functional response in nerve and in striated muscle. According to this law, the response in a single functional unit, although it may vary in size according to the condition of the tissue, cannot be varied in size or character by varying the strength of stimulus. If the stimulus be adequate, the response will be as large as the tissue is capable of yielding at the moment. Gradation of both sensory disturbance and motor response depends on the number of nerve-fibers involved and the frequency of impulses in each fiber. The usual frequencies found in those neural discharges which have been so far investigated lie between 10 and 100 per second.

5) In both nerve and muscle, functional activity involves metabolism, as shown by heat production, consumption of oxygen, and production of CO_2. In this respect there is a great quantitative difference, for the amount of metabolic activity in nerve is so much smaller than in muscle that only the most refined of modern methods have sufficed to measure it. Correspondingly slight is fatigue in nerve; yet it has recently been demonstrated and found to resemble qualitatively fatigue in muscle.

6) Muscular contraction involves the transformation of chemical into mechanical energy. Glycogen is converted into lactic acid with development of tension and liberation of heat. During recovery, part of the lactic acid is oxidized. Part of the resulting energy is utilized in converting the rest of the lactic acid into glycogen, and the remainder of the energy appears as heat.

7) The functions of the nerve-centers, in which they differ from the conducting axons in peripheral nerves, are typified in the spinal reflexes. Numerous evident differences have been described by investigators, notably, in the case of the reflex, greater fatigability and susceptibility to drugs, irreversible conduction, and the presence of after-discharge, summation, and inhibition.

8) Analysis of reflexes has failed to reveal in them any form of behavior which necessarily demands the assumption of a different type of conduction from that which is found in peripheral nerve. On the other hand, Sherrington has suggested the presence in the centers of hypothetical excitatory and inhibitory states or substances which may be graded in amount, and do not involve a refractory phase. In some respects this view of reflex functions is simpler in that it obviates the need of certain subsidiary

hypotheses which would be needed in order to interpret reflexes on the basis of nerve-impulses of the all-or-none type throughout the conducting paths.

9) Tonus is a state of sustained contractile tension in a muscle. It has been believed by some to involve a different kind of function from contraction, but most of the evidence indicates that there is no essential qualitative difference between tonic and other forms of contraction.

10) Integration in the central nervous system, whereby the impulses in the constituent neurons are made the vehicle of the higher activities of man, is a matter of arrangement and interaction. It is not necessary to assume that it involves a "vital force."

REFERENCES

1. Adrian, E. D. On the conduction of subnormal disturbances in normal nerve. *J. Physiol.*, 1912, **45**, 389-412.
2. ———. The all-or-none principle in nerve. *J. Physiol.*, 1914, **47**, 460-474.
3. ———. The recovery process of excitable tissues. Part I. *J. Physiol.*, 1920, **54**, 1-31.
4. ———. The recovery process of excitable tissues. Part II. *J. Physiol.*, 1921, **55**, 193-225.
5. ———. The relation between the stimulus and the electric response in a single muscle fibre. *Arch. neérl. de physiol.*, 1922, **7**, 330-332.
6. ———. Some recent work on inhibition. *Brain*, 1924, **47**, 399-416.
7. ———. The impulses produced by sensory nerve-endings. Part I. *J. Physiol.*, 1926, **61**, 49-72.
8. ———. The impulses produced by sensory nerve-endings. Part IV. Impulses from pain receptors. *J. Physiol.*, 1926, **62**, 33-51.
9. ———. The basis of sensation. New York: Norton, 1928.
10. Adrian, E. D., & Bronk, D. W. The discharge of impulses in motor nerve fibres. Part I. Impulses in single fibres of the phrenic nerve. *J. Physiol.*, 1928, **66**, 81-101.
11. ———. The discharge of impulses in motor nerve fibres. Part II. The frequency of discharge in reflex and voluntary contractions. *J. Physiol.*, 1929, **67**, 119-151.
12. Adrian, E. D., & Lucas, K. On the summation of propagated disturbances in nerve and muscle. *J. Physiol.*, 1912, **44**, 68-124.
13. Adrian, E. D., & Zotterman, Y. The impulses produced by sensory nerve-endings. Part II. The response of a single end-organ. *J. Physiol.*, 1926, **61**, 151-171.
14. ———. Impulses produced by sensory nerve-endings. Part III. Impulses set up by touch and pressure. *J. Physiol.*, 1926, **61**, 465-483.
15 Bernstein, J. Untersuchungen zur Thermodynamik der bioelektrischen Ströme. *Pflüg. Arch. f. d. ges. Physiol.*, 1902, **92**, 521-562.
16. Bishop, G. H. The form of the record of the action potential of vertebrate nerve at the stimulated region. *Amer. J. Physiol.*, 1927, **82**, 462-477.
17. Bishop, G. H., & Erlanger, J. The effects of polarization upon the activity of vertebrate nerve. *Amer. J. Physiol.*, 1926, **78**, 630-657.
18. Bramwell, J. C. B., & Lucas, K. On the relation of the refractory period to the propagated disturbance in nerve. *J. Physiol.*, 1911, **42**, 495-511.
19. Camis, M. On the unity of motor centres. *J. Physiol.*, 1909, **39**, 228-234.
20. Child, C. M., Physiological foundations of behavior. New York: Holt, 1924.

21. Cobb, S. Review on the tonus of skeletal muscle. *Physiol. Revs.*, 1925, **5**, 518-550.

22. ————. Physiology, psychiatry, and the inhibitions. *Arch. Neur. & Psychiat.*, 1928, **19**, 981-996.

23. Cooper, S. The conduction of a nervous impulse in a narcotized region of nerve. *J. Physiol.*, 1926, **59**, 305-318.

24. Cooper, S., Denny-Brown, D. E., & Sherrington, C. Reflex fractionation of a muscle. *Proc. Roy. Soc. (B)*, 1926, **100**, 448-462.

25. Davis, H. The conduction of the nerve impulse. *Physiol. Revs.*, 1926, **6**, 547-595.

26. Davis, H., Forbes, A., Brunswick, D., & Hopkins, A. McH. Studies of the nerve impulse. II. The question of decrement. *Amer. J. Physiol.*, 1926, **76**, 448-471.

27. Davis, H., Pascual, W., & Rice, L. H. Quantitative studies of the nerve impulse. III. The effect of carbon dioxide on the action current of medullated nerve. *Amer. J. Physiol.*, 1928, **86**, 706-724.

28. Downing, A. C., Gerard, W. R., & Hill, A. V. The heat production of nerve. *Proc. Roy. Soc. (B)*, 1926, **100**, 223-251.

29. Erlanger, J. Analysis of the action potential in nerve. Harvey Lectures, Series 22, 90-113. Philadelphia: Lippincott, 1926-27.

30. Fenn, W. O. Gas exchange of nerve during stimulation. *Amer. J. Physiol.*, 1927, **80**, 327-346.

31. ————. Oxygen consumption of frog nerve during stimulation. *J. Gen. Physiol.*, 1927, **10**, 767-779.

32. Field, H., & Brücke, E. Th. Ueber die Dauer des Refraktärstadiums des Nervens bei Ermüdung und Erholung. *Pflüg. Arch. f. d. ges. Physiol.*, 1926, **214**, 103-111.

33. Forbes, A. The place of incidence of reflex fatigue. *Amer. J. Physiol.*, 1912, **31**, 102-124.

34. ————. The modification of the crossed extension reflex by light etherization and its bearing on the dual nature of spinal reflex innervation. *Amer. J. Physiol.*, 1921, **56**, 273-312.

35. ————. The interpretation of spinal reflexes in terms of present knowledge of nerve conduction. *Physiol. Revs.*, 1922, **2**, 361-414.

36. ————. Spiritualism and telepathy. *Harvard Grad. Mag.*, 1927, **36**, 189-202.

37. ————. The problem of tonus in skeletal muscle in relation to sympathetic innervation. *Arch. Neur. & Psychiat.*, 1929. (In press.)

38. Forbes, A., & Cattell, McK. Electrical studies in mammalian reflexes. IV. The crossed extension reflex. *Amer. J. Physiol.*, 1924, **70**, 140-173.

39. Forbes, A., Cobb, S., & Cattell, H. Electrical studies in mammalian reflexes. III. Immediate changes in the flexion reflex after spinal transection. *Amer. J. Physiol.*, 1923, **65**, 30-44.

40. Forbes, A., & Gregg, A. Electrical studies in mammalian reflexes. I. The flexion reflex. *Amer. J. Physiol.*, 1915, **37**, 118-176.

41. ————. Electrical studies in mammalian reflexes. II. The correlation between strength of stimuli and the direct and reflex nerve response. *Amer. J. Physiol.*, 1915, **39**, 172-235.

42. Forbes, A., & Miller, R. H. The effect of ether anesthesia on afferent paths in the decerebrate animal. *Amer. J. Physiol.*, 1922, **62**, 113-139.

43. Forbes, A., & Olmsted, J. M. D. The frequency of motor nerve impulses in the crossed extension reflex as shown by the alcohol block method. *Amer. J. Physiol.*, 1925, **73**, 17-62.

44. Forbes, A., Ray, L. H., & Griffith, F. R. Jr. The nature of the delay in

the response to the second of two stimuli in nerve and in the nerve-muscle preparation. *Amer. J. Physiol.,* 1923, **66**, 553-617.

44*a*. Forbes, A., & Rice, L. H. Quantitative studies of the nerve impulses. IV. Fatique in peripheral nerve. *Amer. J. Physiol.,* 1929, **90**, 119-145.

45. Forbes, A., Whitaker, L. R., & Fulton, J. F. The effect of reflex excitation and inhibition on the response of a muscle to stimulation through its motor nerve. *Amer. J. Physiol.,* 1927, **82**, 693-716.

46. Fulton, J. F. Muscular contraction and the reflex control of movement. Baltimore: Williams & Wilkins, 1926.

47. Gasser, H. S. The relation of the shape of the action potential of nerve to conduction velocity. *Amer. J. Physiol.,* 1928, **84**, 699-711.

48. Gasser, H. S., & Erlanger, J. A study of the action currents of nerve with the cathode ray oscillograph. *Amer. J. Physiol.,* 1922, **62**, 496-524.

49. Gerard, R. W. The two phases of heat production of nerve. *J. Physiol.,* 1927, **62**, 349-363.

50. ———. Studies on nerve metabolism. I. The influence of oxygen lack on heat production and action current. *J. Physiol.,* 1927, **63**, 280-298.

51. ———. Studies on nerve metabolism. II. Respiration in oxygen and nitrogen. *Amer. J. Physiol.,* 1927, **82**, 381-404.

52. ———. The activity of nerve. *Science,* 1927, **66**, 495-499.

53. Gerard, R. W., & Forbes, A. "Fatigue" of the flexion reflex, and a note on action currents and "equilibration" in the cat's peroneal nerve. *Amer. J. Physiol.,* 1928, **86**, 178-205.

54. Gerard, R. W., Hill, A. V., & Zotterman, Y. Effect of frequency of stimulation on heat production of nerve. *J. Physiol.,* 1927, **63**, 130-143.

55. Gotch, F. The submaximal electric response of nerve to a single stimulus. *J. Physiol.,* 1902, **28**, 395-416.

56. Hill, A. V. A new mathematical treatment of changes of ionic concentration in muscle and nerve under the action of electric currents, with a theory as to their mode of excitation. *J. Physiol.,* 1910, **40**, 190-224.

57. ———. The position occupied by the production of heat, in the chain of processes constituting a muscular contraction. *J. Physiol.,* 1911, **42**, 1-43.

58. ———. The absolute mechanical efficiency of the contraction of an isolated muscle. *J. Physiol.,* 1913, **46**, 435-469.

59. ———. The four phases of heat-production of muscle. *J. Physiol.,* 1920, **54**, 84-128.

60. ———. The laws of muscular motion. *Proc. Roy. Soc.,* (*B*), 1926, **100**, 87-108.

61. ———. Muscular activity. Baltimore: Williams & Wilkins, 1926.

62. ———. Living machinery. New York: Harcourt, Brace, 1927.

63. ———. The rôle of oxidation in maintaining the dynamic equilibrium of the muscle cell. The absolute value of the isometric heat coefficient TL/H in a muscle twitch, and the effect of stimulation and fatigue. The absence of delayed anaërobic heat in a series of muscle twitches. The recovery heat-production in oxygen after a series of muscle twitches. *Proc. Roy. Soc.* (*B*), 1928, **103**, 138-191.

64. Hinsey, J. C. Some observations on the innervation of skeletal muscle of the cat. *J. Comp. Neur.,* 1927, **44**, 87-195.

65. Howell, W. H. Inhibition. *Physiol. Revs.,* 1925, **5**, 161-181.

66. Jennings, H. S. The psychology of a protozoan. *Amer. J. Psychol.,* 1899, **10**, 503-515.

67. Jolly, W. A. On the time relations of the knee-jerk and simple reflexes. *Quar. J. Exper. Physiol.,* 1911, **4**, 67-87.

68. Kato, G. The theory of decrementless conduction in narcotised region of nerve. Tokyo, 1924.

69. ————. The further studies on decrementless conduction. Tokyo, 1926.

70. Lapicque, L. L'excitabilité eu fonction du temps. La chronaxie, sa signification et sa mesure. Paris: Les Presses Universitaires de France, 1926.

71. Lee, F. S., & Everingham, S. Pseudo-fatigue of the spinal cord. *Amer. J. Physiol.,* 1909, **24**, 384-390.

72. Lennox, W. G., & Cobb, S. Epilepsy. *Med. Monog.,* 1928, **14**.

73. Levin, A. Fatigue, retention of action current and recovery in crustacean nerve. *J. Physiol.,* 1927, **63**, 113-129.

74. Liddell, E. G. T., & Sherrington, C. S. A comparison between certain features of the spinal flexor reflex and of the decerebrate extensor refiex respectively. *Proc. Roy. Soc.* (*B*), 1923, **95**, 299-339.

75. ————. Recruitment type of reflexes. *Proc. Roy. Soc.* (*B*), 1923, **95**, 407-412.

76. Lillie, R. S. The conditions determining the rate of conduction in irritable tissues and especially nerve. *Amer. J. Physiol.,* 1914, **34**, 414-445.

77. ————. The nature of protoplasmic and nervous transmission. *J. Phys. Chem.,* 1920, **24**, 165-191.

78. ————. Transmission of physiological influence in protoplasmic systems, especially nerve. *Physiol. Revs.,* 1922, **2**, 1-37.

79. Loeb, J. Comparative physiology of the brain and comparative psychology. New York: Putnam, 1903.

80. Lucas, K. The "all or none" contraction of the amphibian skeletal muscle fibre. *J. Physiol.,* 1909, **38**, 113-133.

81. ————. An analysis of changes and differences in the excitatory process of nerves and muscles based on the physical theory of excitation. *J. Physiol.,* 1910, **40**, 225-249.

82. ————. On the transference of the propagated disturbance from nerve to muscle, with especial reference to the apparent inhibition described by Wedensky. *J. Physiol.,* 1911, **43**, 46-90.

83. ————. The conduction of the nervous impulse. London: Longmans, 1917.

84. Lutz, B. R. Threshold values in the spinal frog. II. Variations with change of temperature. *Amer. J. Physiol.,* 1918, **45**, 515-527.

85. MacCurdy, J. T. Common principles in psychology and physiology. Cambridge: Cambridge Univ. Press, 1928.

86. Magnus, R. Zur Regelung der Bewegungen durch das Zentralnervensystem. 3 and 4 Mitt. *Pflüg. Arch. f. d. ges. Physiol.,* 1910, **134**, 545-597.

87. Martin, E. G. The application of the "all-or-nothing" principle of nervous conduction to the interpretation of vasomotor reflexes. *Amer. J. Physiol.,* 1922, **59**, 400-412.

88. Morgan, C. L. Introduction to comparative psychology. London: Scott, 1901.

89. Nernst, W. Zur Theorie der elektrischen Reizung. *Gött. Nachr. Math. Phys. Klasse.,* 1899, 104-108.

90. ————. Zur Theorie des elektrischen Reizes. *Pflüg. Arch. f. d. ges. Physiol.,* 1908, **122**, 275-314.

91. Orbeli, L. A. Die sympathische Innervation der Skelettmuskeln. *J. Petrograd Med. Instit.,* 1923, **6**, 8. Abstract in *Med Sci. Absts. & Revs.,* 1924, **10**, 486.

92. Ostwald, W. Elektrische Eigenschaften halbdurchlässiger Scheidewände. *Zsch. Phys. Chem.,* 1890, **6**, 71-82.

93. Parker, G. H. The origin and significance of the primitive nervous system. *Proc. Amer. Phil. Soc.*, 1911, **50**, 217-225.

94. ———. The origin and evolution of the nervous system. *Pop. Sci. Mo.*, 1914, **84**, 118-127.

95. ———. The elementary nervous system. Philadelphia: Lippincott, 1919.

96. ———. The production of carbon dioxide by nerve. *J. Gen. Physiol.*, 1925, **7**, 641-669.

97. ———. The excretion of carbon dioxide by frog nerve. *J. Gen. Physiol.*, 1925, **8**, 21-31.

98. ———. The carbon dioxide excreted in one minute by one centimeter of nerve fiber. *J. Gen. Physiol.*, 1925, **9**, 191-195.

99. Pratt, F. H. The all-or-none principle in graded response of skeletal muscle. *Amer. J. Physiol.*, 1917, **44**, 517-542.

100. Querido, A. The function of the peripheral neurones in the conduction of impulses in the sympathetic nervous system. *Amer. J. Physiol.*, 1924, **70**, 29-57.

101. Rice, L. H., & Davis, H. Uniformity of narcosis in peripheral nerve. *Amer. J. Physiol.*, 1928, **87**, 73-84.

102. Sherrington, C. S. The integrative action of the nervous system. New York: Scribner's, 1906.

103. ———. On reciprocal innervation of antagonistic muscles. Thirteenth note.—On the antagonism between reflex inhibition and reflex excitation. *Proc. Roy. Soc.* (*B*), 1908, **80**, 565-578.

104. ———. Flexion-reflex of the limb, crossed extension-reflex, and reflex stepping and standing. *J. Physiol.*, 1910, **40**, 28-121.

105. ———. Reciprocal innervation of antagonistic muscles. Fourteenth note.—On double reciprocal innervation. *Proc. Roy. Soc.* (*B*), 1910, **81**, 249-268.

106. ———. Reflex inhibition as a factor in the coördination of movements and postures. *Quar. J. Exper. Physiol.*, 1913, **6**, 251-310.

107. ———. Break-shock reflexes and supramaximal contraction-response of mammalian nerve-muscle to single shock stimuli. *Proc. Roy Soc.* (*B*), 1921, **92**, 245-258.

108. ———. Remarks on some aspects of reflex inhibition. *Proc. Roy. Soc.* (*B*), 1925, **97**, 519-545.

109. Sherrington, C. S., & Sowton, S. C. M. Chloroform and reversal of reflex effect. *J. Physiol.*, 1911, **42**, 383-388.

110. ———. Reversal of the reflex effect of an afferent nerve by altering the character of the electrical stimulus applied. *Proc. Roy. Soc.* (*B*), 1911, **83**, 435-446.

111. Stiles, P. G. Types of fatigue. *Amer. J. Pub. Health*, 1920, **10**, 653-656.

112. ———. The nervous system and its conservation. Philadelphia: Saunders, 1924.

113. Tashiro, S. Carbon dioxide production from nerve fibers when resting and when stimulated; a contribution to the chemical basis of irritability. *Amer. J. Physiol.*, 1913, **32**, 107-145.

114. ———. Ueber die physikochemischen Bedingungen der Erregungsleitung in Nerven. II. Der Einfluss der Ionenkonzentration und des osmotischen Druckes. *Zsch. f. Biol.*, 1925, **83**, 399-414.

115. Verworn, M. Zur Kenntniss der physiologischen Wirkungen des Strychnins. *Arch. f. Physiol.*, 1900, 385-414.

116. Wedensky, N. Ueber einige Beziehungen zwischen der Reizstärke und der Tetanushöhe bei indirecter Reizung. *Pflüg. Arch. f. d. ges. Physiol.*, 1885, **37**, 69-72.

CHAPTER 4

VISION: I. VISUAL PHENOMENA AND THEIR STIMULUS CORRELATIONS

L. T. Troland
Harvard University

I. Introductory

The facts and principles of vision can be presented in a number of different ways. First, we may describe visual phenomena from a purely introspective standpoint, without reference to their physical or physiological conditions. Secondly, we can adopt a strictly behavioristic point of view and try to arrive at an entirely objective account of what happens in the optic nervous system, or the response arc as a whole, when certain stimuli act upon the eyes. Thirdly, we may concern ourselves primarily with the *correlations* between the subjective phenomena and the accompanying physical or physiological factors. Finally, we may attempt to explain any of these observed facts by the introduction of hypotheses which are supposed to describe hidden mechanisms.

In the present contribution, I shall confine myself mainly to the first and the third of these methods, although with some incidental applications of the second method. Considerations of a hypothetical character will be left for the ensuing chapter, "The Nature of the Photoreceptor Process," contributed by Dr. Hecht. It is, of course, difficult to draw an exact line of demarcation between empirical and theoretical ideas, but "theories of vision" have been concerned mainly with receptor and optic nerve activities.

II. The Principal Experiential Factors In Vision

In beginning to consider the facts of vision from the introspective standpoint, we must first establish the principal concepts which are involved. The subjective phenomena of vision consist of certain experiences which can be reported, without physical inference, by the given observer under the given circumstances. These phenomena are found by him in a restricted section of his experience, which may be called the *subjective visual field.* This field comprises a three-dimensional spatial manifold lying mainly in front of the empirical eye,[1] although extending to some *degree* backwards on either side. It is immaterial, from the introspective standpoint, whether or not we regard this visual space manifold as being coincident with the corresponding physical space, but its characteristics must be established by methods of introspective description and not by the inferential methods of objective science.

---

[1] The phrase, "the empirical eye," is used to designate the apparent position of the observer with respect to the remainder of visual space, subjectively considered.

The visual field is characterized not only by its special relation to the remainder of experience but by the peculiar qualities and the structural principles of its contents. As a rule, these contents show a high degree of complexity and definiteness of pattern, as compared with appearances in other experiential fields. Under the most common conditions of vision they comprise visual *things* "seen" in experiential space before the eyes. However, they also include "phantastic visual appearances" such as after-images, phosphenes, idioretinal clouds, and the like. In general, these visual patterns possess height, breadth, and thickness, and are located vertically, horizontally, and depthwise, within the visual field.

Apart from such specific forms and positions, they can present only a finite number of qualitative characteristics, most prominent among which are those which we describe by the term *color*. (For a discussion of the reasons for this use of the term *color*, see 65.) Color can be regarded as constituting the substance, as opposed to the form, of visual presentation. However, color is usually, although not exclusively, confined to the *surfaces* of visual things, so that in order to arrive at a complete analytical formulation of visual experience we must also postulate a spatial or *depth* factor. This depth constituent intervenes between the color surface and the empirical eye, and may also contribute to the volume of visual things, behind or within the color superficies.

Visual experience presents a crucial case for the debate concerning the extent to which *analysis* is legitimate in introspective psychology. Reference is made here to the debate between the structuralists and the configurationists in psychology (63). Visual introspection is practically coincident, in many cases, with the common-sense observation of external things or events. It is self-evident that the visual field can be marked off into mutually exclusive sections and that a great many such sections can exist simultaneously in the same field. However, the possibility of subdivision has a limit, corresponding to the minimum visible area, and ordinary visual patterns are not atomic mosaics but comprise sizable homogeneous areas or volumes. When we speak of *elements* of visual experience, we are not concerned necessarily with sensory atoms which have a fixed visual size but rather with the qualitative character of possible visual volumes and surfaces, independent of their specific forms, in any given instance. However, if anything is gained thereby, there would seem to be no objection to treating any homogeneous visual surface or volume as if it were composed of a number of accurately juxtaposed smaller spatial components.

It has been customary, in discussions of vision, to distinguish rather categorically between sensation and perception (64). The word sensation has ordinarily been used in visual psychology to stand for subjective elements or substances, together with their manner of determination by stimulus characteristics. The term perception on the other hand, has been applied mainly to the form or structure of the experience either in space or in time. It is well to distinguish between these two aspects of "form" and "matter," but little is gained by following the traditional terminology of sensation and perception.

In addition to the static phenomena of vision, we must also consider *changes* or processes which occur within the visual field. These may be alterations either in quality or in form. As an example of the former we may mention "flicker," and of the latter, motion.

The Properties of Color. As indicated above, the elementary constituents of visual experience consist of colors and depth components. The term *color* has been used loosely by physicists to stand for a supposed property of objects or of radiant energy, and by the layman to designate the *chromatic,* as opposed to the *achromatic,* aspects of visual things. However, the preferred usage of the term makes it apply to the purely experiential qualities of visual presentations in general. Color, in this sense, includes black, white, and the grays, as well as visual qualities which manifest hue. The full meaning of the term can be demonstrated best by an enumeration and classification of the different color species, to which we shall now proceed.

The psychological color solid. Colors can be classified in a number of different ways, but the most fundamental division is undoubtedly that between achromatic and chromatic colors. The former can be arranged into a linear or unidimensional series, running from black to white through a continuum of grays. In such a series, comprising all possible achromatic colors, adjacent members will be "just-noticeably different" from each other. This scale of colors establishes a dimension of qualitative variation, which is best designated by the term *brilliance* (65, p. 534). The brilliance value of any color can be specified numerically by counting the number of just-noticeable steps which intervene between it and *black.* In practice, difficulty is found in identifying absolute black, so that brilliance measures are usually relative to some arbitrarily chosen dark color. It would also be feasible to measure brilliance in terms of negative, as well as of positive, values with respect to a mid-gray (*vide infra*).

The *chromatic colors* consist of all colors in addition to the achromatic ones. They are characterized by the possession of a definite *hue*: red, yellow, green, blue, or intermediates between adjacent pairs of these hues. If we attempt to arrange the chromatic colors into a continuum, we find that the latter must be three-dimensional. One of these dimensions corresponds to that of brilliance in the achromatic series. Moreover, we find that a place is automatically reserved in the chromatic solid for the achromatic system as a whole, so that the latter becomes a sort of axis or core of the former, parallel to the purely brilliance dimension. Variations perpendicular to this axis represent the degree of departure of any chromatic color from the achromatic one of the same brilliance. Such variations are in the dimension known as *saturation.* If we treat saturation values as radial, then what we call *hue* corresponds to the tangential or circumferential dimension. Saturation can be regarded as a measure of the degree or intensity of the given hue. These relations are exhibited in Figure 1.

The system of colors, thus established, is characterized by its dimensions rather than by its boundaries. The latter are theoretically indeterminate but, in practice, are limited by physiological conditions. The familiar construction of the color solid as a double pyramid (64, p. 63; for a recent discussion, see 15) is warranted neither by the fundamental theory nor by the psychophysical facts. However, it should be recognized that variation along the hue dimension is characterized by a peculiar kind of discontinuity, which can be expressed by saying that the *mode of variation* changes suddenly at four points along the hue cycle. These points identify the four psychologically primary hues: red, yellow, green, and blue (36).

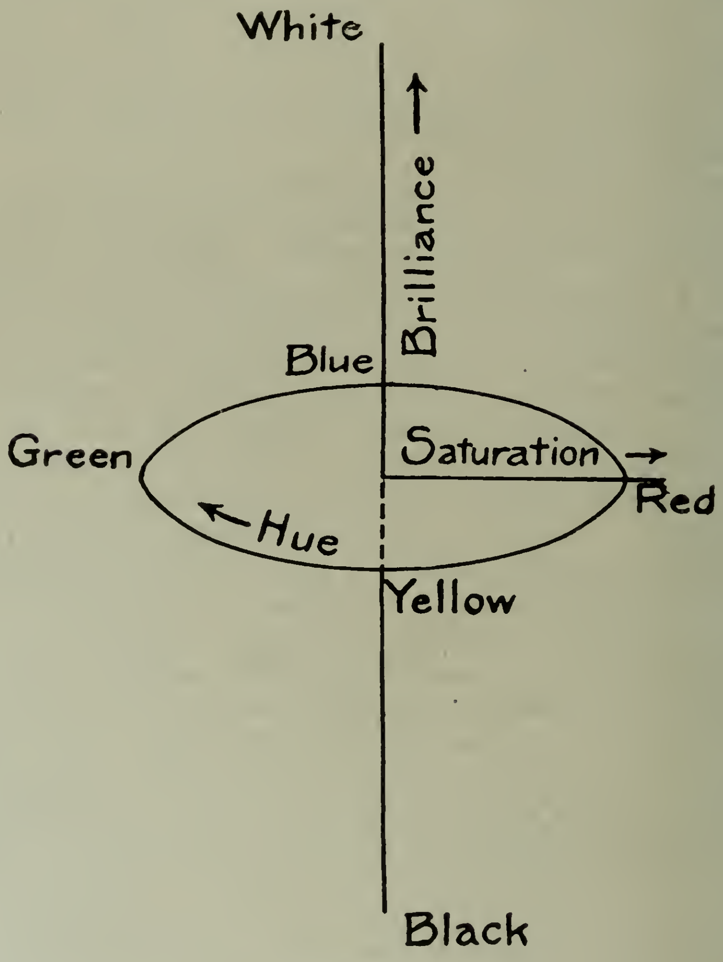


FIGURE 1

THE DIMENSIONS OF THE PSYCHOLOGICAL COLOR SOLID

(From: Report of the Colorimetry Committee of the Optical Society of America, L. T. Troland, Chairman. *J. Opt. Soc. Amer.*, 1922, 6.)

The problem of providing numerical designations for the chromatic attributes, hue and saturation, involves counting the number of just-noticeable differences, in each of these respects, from some standard reference point or member. For saturation, the natural reference point is the achromatic color of the same brilliance, the saturation steps being traced centrifugally along the chromatic radii. In the designation of hue, one of the psychological primaries may be chosen as a zero point. For this purpose, the red is probably the most convenient, the hue being traced circumferentially along a line of constant saturation and brilliance from the red through the yellow, and so on.

It has been claimed by Dimmick (14) that the neutral mid-gray can be recognized introspectively with a facility equal to that which applies to the psychologically primary hues. On this basis, he proposes to make the mid-gray a general reference point for the entire color space, treating black and white as hues. We can then speak of the saturation of black or white and can regard the whole system as being spherically symmetrical about the mid-gray point. However, this scheme does not get away from the three-dimensionality of the manifold, and the plan is of questionable significance on account of the evident dissimilarity between the achromatic and any of the chromatic lines through the mid-gray.

Modes of color appearance. The classification of colors in terms of brilliance, hue, and saturation is by no means exhaustive. Under special conditions, colors present other aspects which have been called "modes of appearance" (*Erscheinungsweisen*). These conditions can be specified introspectively as well as objectively. In our everyday experience we have to deal mainly with *surface colors,* which seem to be confined to, and to constitute, the superficies of visual things. Such surface colors are ordinarily conditional upon a "full" visual field, showing a complex pattern of visual things. When, however, the major portion of the visual field is dark and blank, with an isolated patch of color somewhere in its extent, this patch usually takes the form of a so-called *film,* which is the common experience in connection with spectroscopic apparatus. The experience of a transparent colored object, such as glass, frequently presents a volume or *bulky* color. Self-luminous things, recognized as such, have a peculiar aspect known as *glow,* while metallic and other regularly reflecting bodies manifest the attribute of *luster.* Still further modes of color appearance have been signalized by the work of Katz (33), Martin (41), and others.

It should be noted that the primary attributes of colors are not independent, for their realization or specific values, of the mode of appearance. Thus, good blacks and other dark colors, such as browns and olives, are found only as surface colors or in a surface-color environment. They demand a generally brilliant pattern of outlying things. The general set-up or Gestalt of the visual world has an important bearing upon the appearance of colors. Under the conditions for surface colors there is ordinarily an "abstraction" (*Berücksichtigung*) of the apparent illumination (30), the latter being presented to some extent as a brilliance value which is independent of that attaching to any given surface color. Atmospheric haze is also abstracted in a similar manner, being placed in the space between the thing and the empirical eye (8). These effects are sometimes summarized in terms of a principle of the "color constancy of visual things" (23), according to which surface and volume colors tend to be independent of the illumination and atmospheric conditions.

General Properties of Visual Space. Visual space can be characterized, in general, by stating that it has a *perspective* structure or presentation to a certain point of view, that of the empirical eye. However, this should not be interpreted as meaning that visual space is actually a mere perspective or a projection upon a plane, since it is definitely given as a volume. The visual space field can be described as being approximately hemispherical in form, fading off into non-existence perpendicularly to the line of sight in all directions, and disappearing by a process of condensation along the latter line (39). In general, only a single color surface is presented along any one line of view, in accordance with the principle of *visual eclipse.* A few exceptions to this rule appear in cases of transparent objects where one surface may be presented directly behind another.

As we have already indicated above, normal visual space can be described analytically in terms of a sheaf of lines of depth elements radially arranged with respect to the empirical eye. Each one of these lines ordinarily terminates upon an elementary color surface. By an effort of abstraction, common to artists, any given pattern in the visual field can be viewed as projected upon a single plane, perpendicular to the empirical line of sight. Such a plane may be designated as a *subjective projection field.* It is of importance in separating the influences of monocular from those of binocular vision.

General Plan of the Ensuing Discussion. We may now turn our attention to the psychophysical relationships of the various subjective factors which we have just considered. The psychophysical treatment will bring out further details of the introspective phenomena, in the appropriate contexts. It will be convenient to divide this treatment into a number of separate sections, dealing respectively with (1) the nature of the visual stimulus, (2) stimulus correlations of brilliance, (3) stimulus correlations of chroma, (4) relations of color to binocular factors, (5) monocularly determined form, (6) binocularly determined form, (7) visual motion perception, and (8) a few appended comments on the relation of visual experience to cerebral processes. Although the latter topic does not belong properly under the title assigned to the present chapter, it is added to complete the total discussion of vision comprised by the two chapters, since Dr. Hecht's contribution treats exclusively of receptoral and optic nerve correlations.

III. Characteristics of the Visual Stimulus and of the Eye

Radiant Energy, the Stimulus. The "adequate" visual stimulus consists of radiant energy (12), having a certain limited range of wave-lengths, approximately 400 to 780 mμ. This energy ordinarily emanates from objects outside of the eyes, passing through the pupils and the various ocular media, to form more or less distinct images upon the two retinas. The direct stimulating action is located, of course, in the receptor elements of the latter, but most of our knowledge concerning the correlations between visual phenomena and the stimulus has to do with extra-ocular conditions. However, some important relationships can be traced, deductively, for the intra-ocular stimulus.

The physical concept of radiant energy appears to be a transitional stage at the present day (3). It is agreed that this energy is a form of electromagnetic undulation, capable of travelling through empty space at a tremendous velocity, but its exact mode of distribution within this space is doubtful. Certain phenomena appear to call for a bullet or arrow-like structure of the disturbance (according to the quantum hypothesis), while other effects seem to demand a much more extensive spatial continuity. However, it is fortunate that this theoretical difficulty has very little bearing upon the formulation of known visual facts, which can be treated adequately in terms of the original wave theory of Huyghens and Young (49) or on the basis of experimental measures, apart from any hypothesis.

The general characteristics of any given sample of radiant energy can be specified by means of a so-called *distribution* curve, such as those given in Figure 2. This shows how the intensities of the spectral components of the sample vary as a function of the wave-length. Wave-lengths are ordinarily measured in terms of millionths of a millimeter, or millimicrons, correctly symbolized by mμ. The intensity measure can be adjusted to the requirements of the given situation (66), but is founded upon the unit of energy, the erg. Since a *flow* of energy is ordinarily involved, we usually have to deal with *ergs per second,* and, if we wish to discount the effects of area, the intensity may require expression in terms of *ergs per second per square centimeter.*

Photometric measures. In perhaps the majority of cases, *photometric* intensity measures are more useful than are *radiometric* ones, although it should never be forgotten that the former incorporate a psychophysical criterion. However, we are forced to acknowledge, in any case, that a stimulus must be defined, at least in part, in terms of its power to stimulate the specified sensory system. Photometric measures, or *light* intensity values, presuppose a standard "visibility" factor which represents the capacity of the given form of radiant energy to evoke certain visual effects. The fundamental unit of photometric intensity is the *lumen,* corresponding to a flux of energy. The most useful photometric measure in visual studies (9) is the brightness value of the object surface, which may be expressed in millilamberts or in candles per square meter. In practice this should be multiplied by a factor proportional to the size of the ocular pupil under the given experimental conditions to obtain a measure which is proportional to the intensity of retinal illumination. I have suggested a measure of this sort based upon candles per square meter and pupillary areas expressed in square millimeters yielding a unit designated as the *photon* (66, pp. 24-32).

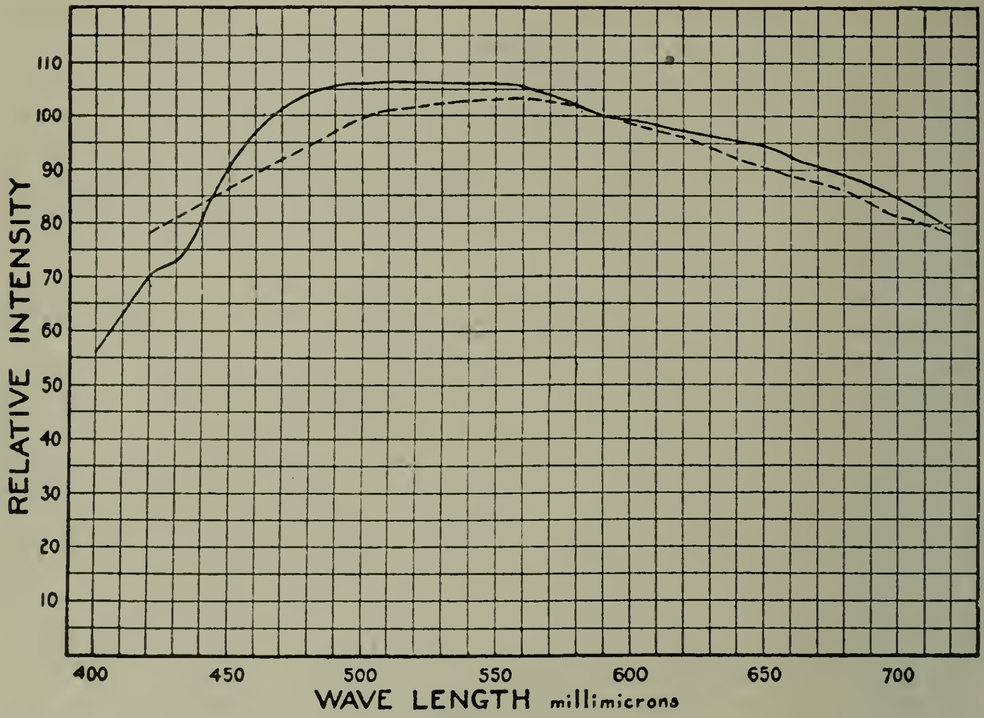


FIGURE 2

Spectral Energy Distributions of Natural and Artificial Sunlight

The solid line represents the distribution of average noon sunlight, while the broken line is that of Priest's precision artificial sunlight. (From: Report of the Colorimetry Committee of the Optical Society of America, L. T. Troland, Chairman. *J. Opt. Soc. Amer.*, 1922, **6**.)

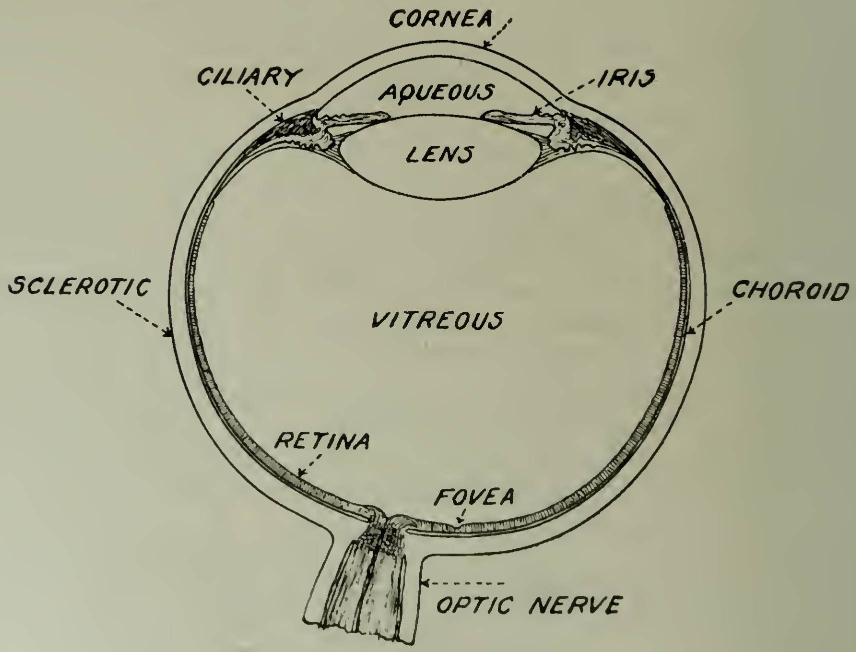


FIGURE 3

Cross-Section of the Human Eye

The figure represents the left eye, as seen from above. (After Quain. To be used in: Troland, L. T. The principles of psychophysiology. New York: Van Nostrand.)

Stimulus attributes. It is usually stated that the visual stimulus has three independent attributes: wave-length, intensity, and purity; but it is evident that the physical association between these three features is very intimate. Thus, changes in purity involve an alteration in wave-length composition, sometimes introducing wave-lengths which were not previously present at all. However, in the case of stimuli having substantially single wave-length constitution (homogeneous stimuli), it will be convenient to consider intensity and wave-length as separate variables. Stimuli having a complex wave-length composition must be treated in accordance with a more complex method.

Stimulus patterns. The above considerations apply primarily to spatially elementary visual stimuli or to very limited pencils of radiant energy entering the eye. The stimulus for the whole eye, however, comprises a definite arrangement of various kinds of radiant energy in space, characterized in any concrete instance by a specific areal or angular pattern. This pattern can be specified by assigning the correct intensity distribution curve to each point in the visual object field or to each direction of the rays which enter the pupils. It is evident that, in general, this configuration of distributions will be slightly different between the right and left eyes, a fact which furnishes the basis for binocularly determined discriminations.

Gross Anatomy and Physiology of the Eye. As shown in Figure 3, the human eye is approximately spherical in form, and its general structure closely resembles that of a photographic camera. The shell of the eyeball is built up of three concentric layers (53). The outermost or sclerotic layer is primarily a supporting tissue but is continuous, in front, with the transparent cornea, the surface of which is responsible for most of the optic refraction. Within and concentric with the sclerotic is found the choroid coat, extended forward to form the iris, the variable opening in which constitutes the pupil, controlling the amount of energy that enters the eye. The third layer comprises the retina or sensitive nervous film. The angular extent of the retina is about 207 degrees. The sensitive layer is perforated, about 18 degrees off center to the nasal side, by the exiting fibers of the optic nerve, and, approximately in the retinal center, there is a small depression, known as the *fovea centralis,* in which the receptor cells are crowded very closely together. Microscopic examination shows that the retina (53, pp. 220-252) consists of three neurologically distinct strata, which, in centripetal order, are: (1) the receptor cells, (2) the bipolar neurons, and (3) the ganglion cells with their attached axons, the latter being continuous with the conducting fibers of the optic nerve.

The rods and cones. The receptor stratum of the retina contains two histologically different species of cells, known as rod and cone cells, respectively (see Figure 4). Both types of cells have three segments, of which the outermost or terminal one is supposed to be the seat of sensory excitation. The terminal segments of the rod cells are rod-shaped, whereas those of the cone cells are conical in form, except in the center of the retina (*fovea centralis*), where they are packed closely together and are elongated so that they closely resemble the rods. Over a three-degree central area there are no true rods. The ratio of rods to cones increases from center to periphery, but there is no region which is free from cones. The central area of the retina also incorporates a layer of yellow pigment, forming the so-called *macula lutea.*

The dioptric process. The function of the eye as an optical instrument consists in forming an image of the external object field upon the retina. This is accomplished by refraction at the outer surface of the cornea and at both surfaces of the crystalline lens. The latter lies directly behind the iris. The refractive power of the cornea is about 43 dioptries, and, of the lens, approximately 21 dioptries. The process of *accommodation,* or of focusing upon objects at different distances from the eye, involves a change in the curvature of the front surface of the crystalline lens, which (in the 10-year-old child) amounts to a 14-dioptry maxi-

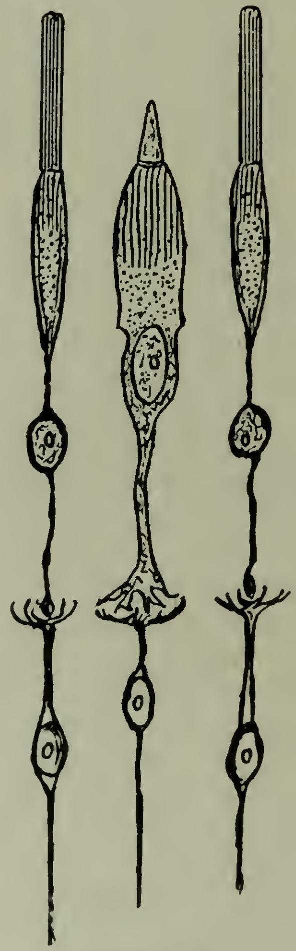


FIGURE 4
RETINAL RODS AND CONES WITH BIPOLAR CELLS
(From: Troland, L. T. The mystery of mind. New York: Van Nostrand, 1926.)

mal alteration in the refractive power. According to Helmholtz' (22) theory, which is now generally accepted, accommodation for near objects is due to a contraction of the ciliary muscles which surround the lens. This contraction reduces the tension of the suspensory ligament, and permits the lens to bulge by virtue of its inherent elasticity.

The chambers between the cornea and the lens, and between the lens and the retina, are filled with transparent substances, the aqueous and the vitreous humors, respectively. The pupil, which is formed on the anterior surface of the lens, has a diameter varying between two and eight millimeters, in accordance with the intensity of the illumination, the state of adaptation of the retina, etc. Its average size is about four millimeters. The laws governing its changes as a function of stimulus intensity, time, etc., have been determined accurately by Reeves (54). The photographic "speed" of the eye, with a maximal pupil, is about F. 2.9.

The form of the image, which is thrown upon the retina by the dioptric mechanism, can be constructed from a knowledge of the object field pattern, in accordance with the usual principles of geometric optics. The two nodal points of the ocular system are very close together so that, for practical purposes, the object or ray pattern can be projected through a single point, located in the posterior portion of the lens, and the extended projection lines, where they strike the retina, will determine the image pattern. Such a projection line, passing through the nodal point and the center of the *fovea centralis,* is called the *line of regard,* and the point in the object field which it intersects is said to be *fixated.* Under normal conditions of vision, the process of binocular convergence brings the lines of regard of the two eyes together at the objective point of fixation. The plane which is established by these two lines is known as the plane of regard (22, vol. 3, pp. 40-43).

It is evident that the size of the retinal image, corresponding to a given object, changes with the distance as well as with the intrinsic dimensions of the object. However, the retinal image size is practically constant for a fixed *angle,* subtended at the ocular nodal point. Consequently, sizes and positions of areas, lines, or points in the object field can be specified most usefully in angular units.

Optical defects of the eye. The refractive action of the eye is subject to most of the aberrations which are characteristic of uncorrected lens systems. (A quantitative study of certain aberrations of the eye is given in 2). These include the spherical, the chromatic, and the astigmatic types. The images of objects upon which the eye is not accommodated will be blurred. There are many different forms of ray-scattering in the eye, as a result of irregularities in the ocular media, and these may produce, upon the retina, characteristic patterns of radiant energy, which are not representative of features of the external object field. They are responsible for many so-called "entoptic" phenomena.

IV. The Psychophysics of Brilliance

We may now commence our discussion of the relations which obtain between subjective visual phenomena and various characteristics of the stimulus, by considering the psychophysics of brilliance. We may regard brilliance as an attribute of color which has a zero value at an ideal black and which increases continuously with increasing resemblance of the color to white. It is an attribute not merely of achromatic but also of chromatic colors. Brilliance is commonly conceived as being especially significant of stimulus intensity, but it is actually a function of all stimulus characteristics as well as of numerous physiological variables. In endeavoring to separate these correlations, with different variables, we must regard all of the remaining factors as being held constant at some arbitrary value. In the present chapter, we shall confine ourselves primarily to characteristics of the stimulus, but some reference must necessarily be made to some of the more important physiological factors in order to establish all of the conditions for the specified stimulus relationships.

Brilliance as a Function of Wave-Length. The most fundamental relationship between brilliance and stimulus variables is that to wave-length. This can be studied most simply by use of homogeneous stimuli. The measurement of radiant intensities in strictly energetic terms immediately demonstrates a lack of proportionality between brilliance and photometric measures, on the one hand, and the quantity of radiant energy, on the other hand. The most intensive constituents of certain common forms of radiant energy are quite invisible.

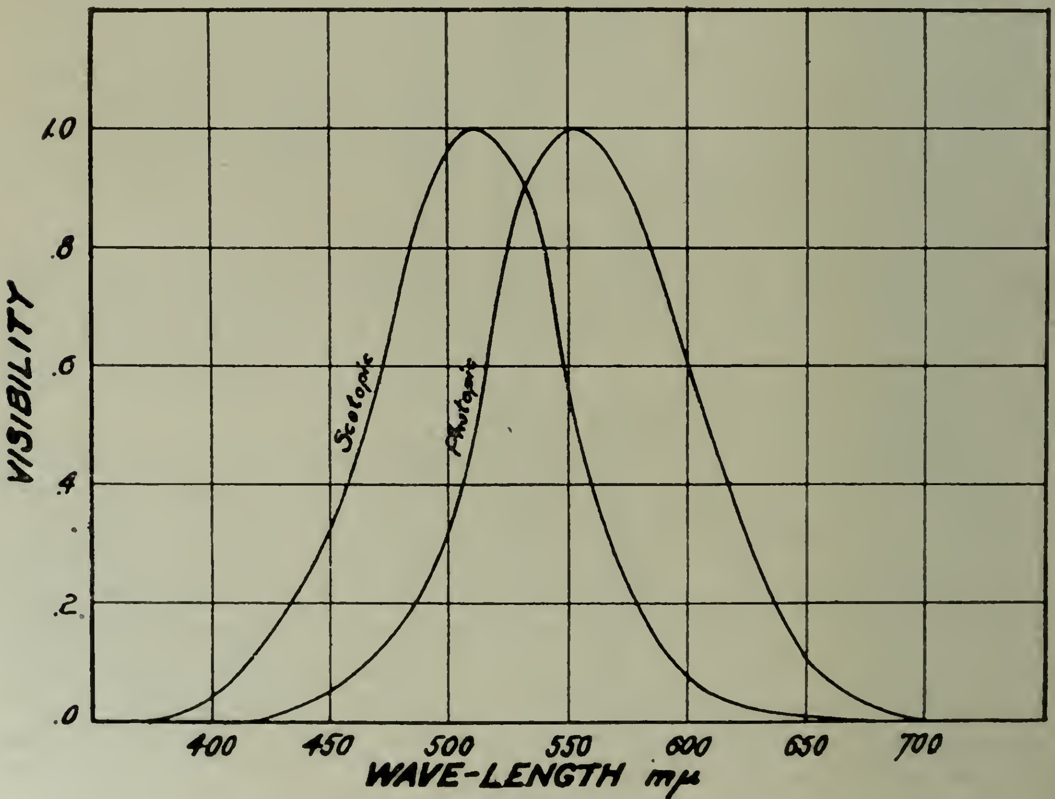


FIGURE 5

The Photopic and Scotopic Visibility Curves of Human Vision

(After Hecht. To be used in: Troland, L. T. The principles of psychophysiology. New York: Van Nostrand.)

Visibility curves. The most thoroughly developed technique for establishing a relationship between brilliance and wave-length is that for the determination of *visibility.* In this procedure, the brilliance is held constant, with varying wave-length, by making the requisite adjustments of intensity. The immediate result is a curve which shows a minimum at about 555 mμ, the values increasing toward unattainable magnitudes at either end of the spectrum. If we compute the reciprocals of these energy values in terms of the minimal value as a unit, we obtain the curve of the so-called *visibility coefficient,* with respect to wave-length, as shown in Figure 5. The corresponding numerical values are presented in Table 1. They are regarded as measures of the relative capacities of radiant energy of different wave-lengths to produce vision.

The exact position of this curve on the wave-length scale varies with the brilliance level at which the measurements are made. It also depends upon the momentary state of adaptation or sensitivity of the retina, although this state is itself determined by the intensity level, if a sufficient time lapse is allowed. Moreover, the curve is affected by abnormal physiological conditions and by congenital visual type, as in the different forms of color-blindness (*vide infra*). It has been shown by Hecht and Williams (21) that, with decreasing intensity, the maximum of the visibility curve shifts from about 554 mμ to approximately 511 mμ, but that subsequently to this shift the general form of the curve remains substantially unchanged. The high intensity curve, or position, is designated as *photopic* and is attributed to the retinal cones, while the low intensity, *scotopic,* curve is assigned to the rods.

This displacement of the region of maximal visibility towards the shorter wave-lengths, with decreasing intensity, is responsible for the so-called *Purkinje phenomenon* and related effects. The Purkinje phenomenon is usually described in terms of a reversal of the brilliance relationships between "red" and "blue" colors, in consequence of diminishing stimulus intensities, without change in their ratio. Even when the red is more brilliant at the higher intensity level, the blue is almost certain to gain superior brilliance at the lower level. It will be seen from the two curves shown in Figure 5 that the extreme long-wave rays of the spectrum have substantially zero visibility at low intensities. Similar changes are noted when a stimulus pattern is moved from the center to the periphery of the visual field, under dark-adaptation at low intensity levels. This is attributed to the fact that rods are absent in the retinal center.

Light. If the brilliance which is evoked by a stimulus of any desired wave-length composition is matched under determinate conditions with that due to radiant energy emitted by a standard candle, a definite relationship can be established between the visibility curve and certain absolute photometric units such as the lumen. This makes it possible to specify an absolute, as opposed to a relative, visibility for each wave-length or wave-length combination. Such absolute visibility is expressed in terms of

TABLE 1

AVERAGE NORMAL VISIBILITY VALUES

Adopted as standard by the Illuminating Engineering Society and the Optical Society of America 1919–1920.

Wave-lengths	Adopted mean I. E. S.	Absolute visibility
400	0.0004	0.27
10	.0012	0.80
20	.0040	2.7
30	.0116	7.72
40	.023	15.3
450	0.038	25.3
60	.060	40.0
70	.091	60.7
80	.139	92.6
90	.208	139
500	0.323	216
10	.484	322
20	.670	446
30	.836	557
40	.942	629
550	0.993	662
60	.996	663
70	.952	641
80	.870	580
90	.757	502
600	0.631	421
10	.503	336
20	.380	253
30	.262	175
40	.170	113
650	0.103	68.8
60	.059	29.3
70	.030	20.0
80	.016	10.7
90	.0081	5.4
700	0.0041	2.7
10	.0021	1.4
20	.0010	0.67
30	.00052	0.35
40	.00025	0.17
750	0.00012	0.08
60	.00006	0.04
Wave-length of maximum	556	

lumens per watt. The high-intensity visibility curve, with a maximum at 554-556 mμ, is employed for this purpose, and the absolute visibility value at the maximum in question is about 617 lumens per watt (11). These relationships permit the photopic visibility curve to be employed in the precise definition of light values, so as to make them independent of a fixed wave-length composition (that of the standard candle) and of accidental variations in photometric matching between individuals.

Direct psychophysical functions. It will be noted that the visibility curves do not establish a strictly psychophysical function connecting brilliance and wave-length. Instead, these curves express a relationship between two stimulus variables which is determined by the subjective criterion of equality-of-brilliance. However, the curve representing brilliance as a function of wave-length, for constant energy, must have a form which is generally similar to that of the visibility relation, although it will be relatively flattened at the peak, and must vary in its exact proportions with the intensity level for which it is computed. It obviously involves a combination of the visibility and the Fechner functions (to be discussed below).

Intra-ocular modifications. It will be noted that the ordinary visibility function is established with respect to the stimulus, as the latter exists outside of the eye. In passing through the ocular media, including the yellow pigment of the retina, the intensity distribution curve of the stimulus is appreciably modified in form. If we apply our knowledge of the selective absorption of the ocular media, we find that the photopic visibility curve, with respect to the stimulus as it must operate upon the retinal receptors, is almost perfectly symmetrical in form, and fits the equation for the probability integral quite closely (67).

Brilliance as a Function of Intensity. The manner in which brilliance varies with intensity can best be discussed under two general headings: (1) the absolute threshold and (2) the Fechner function. Both of these relationships are influenced in an important manner by wave-length composition, as well as by the brilliance of the test-field surroundings. If the intensity of the stimulus is zero over the entire area of the retina, the accompanying experience is not typically a black, but is, instead, a dark gray, which is sometimes known as the "idioretinal light" and is attributed to retinal self-excitation. The lowest values of brilliance appear to be conditional upon a reasonable amount of stimulation of outlying retinal areas (simultaneous contrast) or upon prior excitation of the given area (successive contrast). In discussing these relations, we can abstract from the influence of wave-length by expressing the intensity measures in photometric terms. This is legitimate because the influence in question is essentially of a visibility character.

Absolute threshold for brilliance. If we consider the case of a comparatively small stimulus with a dark outlying field, we find that the absolute threshold varies with position and with the state of adaptation. A three-degree circular spot can be seen in the center of the field at an intensity of about 0.05 photons. In the periphery of the field, with full scotopic dark-adaptation, only about 1/10,000th of this intensity is required. Angularly minute stimuli, such as stars, furnish us with a means for determining the minimal amount of energy which will yield any specific visual effect at all. Studies by a number of investigators (57) show that this minimal value lies in the neighborhood of 10^{-9} ergs per second. If, also, we introduce the minimal time required for vision under these conditions, we find the total needed number of ergs to be approximately 2×10^{-10}, which is not far from the supposed order of magnitude of the ultimate radiation quantum, for radiation of the highest visibility.

The Fechner function. Still assuming a dark outlying field, the mode of variation of brilliance as a function of the intensity, above the threshold, follows a law which we may designate as the Fechner function, without implying that this function is of a strictly logarithmic type. The brilliance itself is measured in terms of the number of just-noticeable changes or steps from a minimal value. In the common form of the experiment, the latter is not a black but is identical with the idioretinal gray. A study of the Fechner function shows that it is approximately logarithmic over a range of medium intensities, the extremes of which stand in a ratio of about one to a hundred. For lower and for higher intensity ranges, the relative threshold increases progressively (45). It may be noted that, although the total intensity scale which can be covered by human vision is from about 0.0001 candles per square meter to a billion candles per square meter, the intensities which we ordinarily use and which yield the highest general visual efficiency lie in a region where the relative threshold is substantially constant or where the logarithmic relationship is practically satisfied. This relationship can be regarded as significant of the tendency of the visual system to record reflection coefficients rather than absolute surface brightnesses. The order of magnitude of the threshold in the optimal region is about one-hundredth part of the total intensity, varying with the exact nature of the discrimination pattern which is involved. It is also to some extent a function of wavelength or wave-length composition, although being to the first approximation independent of these chromatic aspects of the stimulus.

It has been shown by Schjelderup (58) and others that when the surrounding field is not dark, its intensity enters into the Fechner function (or the analogue of this function) in an important way. In general, the just-noticeable difference in intensity is proportional to a linear function of the test and surrounding field values, additively combined. According to Cobb (10) and to Dittmers (16), the minimal value of the threshold is found when the two fields have about the same intensity. For very high values of the surrounding field intensity, the threshold is governed almost exclusively by the latter. It seems likely that at least part of this dependency of the brilliance discrimination upon the brightness of surroundings is due to the scattering of rays within the ocular media, so that an analysis of a similar relationship to the immediate retinal stimulus might yield quite a different result. However, in the case of a completely dark surrounding field, intra-ocular conditions probably influence the function very little.

Brilliance as a Function of Wave-Length Composition. When several stimuli, having different wave-lengths, act upon the same retinal area, the resulting brilliance is determined, at least approximately, by the arithmetic sum of their photometric values (27). If the Fechner function is formulated in terms of photometric intensities, the brilliance will be that which is normal to the summated photometric magnitude. As stated, this principle is valid only for pure photopic vision. In order that it should apply, also, to scotopic and to mixtures of photopic and scotopic vision, the photometric values must be adjusted to the exact visibility curve which is operative under the given conditions.

Brilliance as a Function of Spatial Factors. The brilliance which results from a stimulus of fixed intensity and composition varies with its spatial characteristics and surroundings. These variations involve its size and shape, and the manner of excitation of outlying visual regions.

If a stimulus of fixed internal characteristics is moved from the center to the periphery of the visual field, in general its brilliance will not remain unaltered. The exact law of the change depends upon the composition of the stimulus and upon the state of adaptation of the retina. Under scotopic dark-adaptation, as a rule, the brilliance will increase greatly for all except the extreme long-wave stimuli. Under photopic bright-adaptation, there may actually be a decrease in brilliance towards the periphery. The exact laws which govern such changes have not been determined experimentally, but, as we have seen, absolute thresholds indicate a sensitiveness to "white light" at the periphery, which is about ten thousand times greater than at the center, under full dark-adaptation.

Brilliance and size. When a stimulus varies in size while the intensity is kept constant, the brilliance alters markedly over a range of small sizes but becomes constant at large sizes. These relationships have been studied quantitatively only for threshold conditions, under which it is found that the threshold is inversely proportional to the angular area of the stimulus, up to a diameter of about 50 minutes of arc, in the center of vision. In the periphery, with a purely photopic stimulus, a similar relationship can be demonstrated up to about half a degree (47). Scotopic stimuli, in this retinal region, show a variation of the brilliance threshold inversely as the diameter rather than as the area.

Brilliance contrast. It is a very familiar fact that the brilliance of a given test field varies with the brilliance of outlying areas, in accordance with the principle of contrast. The contrast effect consists either in an increase or a diminution of the brilliance of the test field as a function of the difference in intensity between the two stimulus fields. According to Hess and Pretori (24), and other authorities, the change in the apparent intensity is a linear function of this difference, being positive when the surrounding field is darker and negative when it is brighter than the test field. Within certain limits, if the *intensity changes* of the two fields are proportional, the brilliance of the test field may remain constant. Kirschmann (34) states that the contrast effect is proportional to the square root of the area of the surrounding field. Brilliance contrast is substantially independent of the chromatic characteristics of the experience, if the intensities are expressed in photometric terms.

Brilliance as a Function of Time. When the intensity or other characteristics of the visual stimulus are suddenly changed, the corresponding brilliance value does not alter instantaneously but behaves as if a factor of inertia were involved. However, after the passage of a certain period of time, a new equilibrium is reached. Relations of this sort which are experimentally demonstrable suggest the existence of a number of independent temporally operative factors having quite different values. The inertia is comparatively small in connection with the so-called "persistence of vision," but the time constant is much greater for the phenomena of adaptation.

When a visual stimulus is of short duration—less than one-fifth of a second—the resulting brilliance may depend upon its duration as well as upon its intensity. For any stimulus, a time can be found which is sufficiently short so that the duration and the intensity are interchangeable and the visual process obeys the so-called *reciprocity law.* Under such conditions, the brilliance may be expected to vary as the Fechner function of the *exposure* (intensity times time). Although the reciprocity principle has its limitations in vision (48), as in photochemistry at large, it is nevertheless true that there is no time so brief as to prohibit visual excitation, provided the intensity can be made sufficiently great to bring the exposure value above the threshold.

Growth and decay curves. When a stimulus is instantaneously applied to the retina, the brilliance rises rapidly at first and then more and more slowly until a maximum is reached. The total time occupied by this process is never greater than about one-fifth of a second (43). This duration or *action time* decreases with increasing intensity. When the stimulus is cut off or is suddenly diminished, the brilliance decays in accordance with a curve which is similar to that for its growth but of opposite sign.

These transitions occur so rapidly that they are not ordinarily analyzable by direct introspection. However, when the intensity changes are due to the motion of a sharp image edge across the retina, the brilliance changes are presented as phenomena attached to the edge in question. Experiments by this method show that the first phase of decay of brilliance is usually followed by a recrudescence of the experience in the form of a series of pulses (42). The most rapid of these constitute the so-called *Charpentier's bands.* Slower alterations, succeeding the bands, comprise a series of *positive after-images* of which the most prominent, the Purkinje after-image, is attributed to a delayed scotopic response. The Purkinje after-image is absent or very faint in the center of vision, as would be expected according to this explanation. Under certain conditions—especially with strong dark-adaptation and brief stimulation—very persistent positive after-images of stationary stimuli are obtainable.

Flicker. When two different stimuli are alternated, not too rapidly, there is a corresponding pulsation of the brilliance, known as *flicker.* However, regardless of the magnitude of the difference between the stimuli, a rate of alternation (critical flicker frequency) can always be found at which the flicker disappears. It has been shown that this frequency is proportional to a logarithmic function of the difference in intensity between the two stimuli or of the intensity of a single stimulus which is alternated with a dark interval. In general, this logarithmic law is approximately independent of the chromatic characteristics of the stimulus or experience, if the intensity is expressed in photometric terms. The constants of the law change rather abruptly at a certain low intensity, corresponding to the transition point between photopic and scotopic response, and in the range of lower intensities there is a considerable dependency of the critical frequency upon wave-length features (26).

Under the condition that flicker has been eliminated, the resulting brilliance value is the same as that which would accompany the continuous application of the same average flux of the given kind of radiant energy. This principle, which is known as the Talbot-Plateau law, has been established very accurately, and is a logical corollary of the reciprocity law.

Adaptation. When a stimulus is permitted to act continuously after the brilliance has reached its maximum, the brilliance value ordinarily commences to fall off. This change will be rather rapid at first, gradually decreasing in rate until an equilibrium is reached. A reverse change occurs when the visual system is relieved of a positive stimulus which has been acting for some time. The first type of process has been called *bright-adaptation,* and the second type, *dark-adaptation.* Analysis shows that both of these processes can occur for either photopic or scotopic vision and that the laws for the two kinds of vision are quite different. Dark-adaptation is not ordinarily observable as a change occurring directly in a continuously experienced color, but must be demonstrated by means of brief exposures separated by appropriate time intervals, when the brilliance evoked by a stimulus of fixed intensity can be shown to increase. This points to the conclusion, which is also otherwise indicated, that adaptation stands for a semi-persistent change in the *sensitivity* of the photopic or scotopic visual system, as the case may be.

Photopic adaptation. Photopic adaptation can be demonstrated, to the exclusion of the scotopic variety, by the use of small stimuli—less than three degrees in diameter—restricted to the center of vision. Bright-adaptation is characterized by continuous reduction of the brilliance until an equilibrium value is reached, the time required for this process being about 90 seconds, independently of intensity or wave-length composition, but varying with the area of the test field. The final amount of brilliance reduction is greater, the higher the stimulus intensity, thus tending to level out the differences between the brilliance consequences of different stimuli. However, except when these differences are initially very small, the equilibrium condition does not bring all stimuli to an apparent photometric equivalency (68). The equilibrium sensitivity is inversely proportional to a power of the stimulus intensity which is somewhat less than unity. This sensitivity can be measured by ascertaining the intensity required on the bright-adapted area to match the brilliance evoked by another intensity on an adjacent, fully dark-adapted area, and calculating the ratio between the latter and the former intensity values.

It has been claimed by Abney that photopic bright-adaptation by stimuli of restricted wave-length range reduces the brilliance sensitivity more for the adapting than for other wave-lengths, i. e., that achromatic bright-adaptation is specific with respect to wave-length composition. Attempts made by C. H. Langford and me to verify this alleged effect have been unsuccessful (69), but Frank Allen (1) has demonstrated very interesting selective effects, both for the adapted and the opposite, non-adapted eye, as tested by the method of critical flicker-frequency.

Photopic dark-adaptation occurs very rapidly, but, before it is complete, *negative after-images* can be observed against the idioretinal light. These correspond approximately in form to the primary image of the adapting stimulus but are dark where the latter was light, and vice versa. Such images can also be seen temporarily against a "reacting stimulus" field which is substantially free from detail and is considerably larger than the original adapting stimulus.

Scotopic adaptation. Scotopic dark-adaptation is a much slower process than is the corresponding photopic change. It can be demonstrated, to the exclusion of photopic effects, at intensities lower than the photopic threshold and in a relatively peripheral portion of the visual field. It is responsible for the tremendous increase in the general sensitivity of the eyes under conditions of twilight illumination or in complete darkness. This sensitivity is, of course, characterized spectrally by the dominance of the scotopic visibility curve and by other features associated with response of the retinal rods.

Complete scotopic dark-adaptation requires more than an hour, although the major portion of the change is accomplished in about 20 minutes. The total increase in sensitivity may be by a factor greater than 10,000. However, after a few minutes there is relatively little alteration in the brilliance value elicitable in the center of the visual field because of the absence of rods in the corresponding portion of the retina. Scotopic bright-adaptation occurs much more rapidly than does the dark-adaptative process, and the first phases of such bright-adaptation are accompanied, in bright light, by the experience of *being dazzled.* After brief bright-adaptation, scotopic dark-adaptation occurs more rapidly than when the bright-adaptation is more protracted. The laws relating the sensitivity changes in bright- and dark-adaptation to time are discussed fully in the ensuing chapter by Dr. Hecht. They are of the saturation type, asymptotic in form, and are in harmony with the kinetics of a bimolecular chemical reaction. The brilliance which results from scotopic dark-adaptation is essentially achromatic, corresponding to the failure of the retinal rod system to respond specifically to wave-length.

V. The Psychophysics of the Chromatic Attributes

We may now consider the laws which relate the chromatic attributes, hue and saturation, to stimulus characteristics. These two features of color are so closely associated that they can best be discussed in parallel.

Hue as a Function of Wave-Length. Hue is ordinarily regarded as an index of the wave-length, or wave-length composition, of the stimulus. Although this is by no means the only correlation of hue, it is certainly the most important one. The problem is best considered separately for homogeneous and for heterogeneous stimuli. The spectrum presents an orderly series of substantially single wave-lengths, comprising all possible homogeneous stimuli. Introspection immediately demonstrates the following series of spectral hues, running from the short- to the long-wave end of the spectrum: violet, blue, blue-green, green, yellow-green, yellow, orange, and orange-red. Three psychological primaries are ordinarily localizable with reference to the spectrum, as follows: yellow at 575.5 $m\mu$, green at 505.5 $m\mu$, and blue at 478.5 $m\mu$ (75). The psychologically primary red requires a heterogeneous stimulus, since the extreme red of the spectrum is appreciably yellowish.

The spectral chroma scale. In order to arrive at a more exact formulation of the relation between hue and the physical spectrum, we might propose to step off just-noticeable differences in hue, starting at some arbitrary point, such as the red. Unfortunately, the spectrum yields variations in saturation, as well as in hue, so that we are compelled to describe the results of this experiment as a *spectral chroma scale.* The findings of Jones (31) are given in Table 2. It will be seen that there are 128 discriminable chromas in the spectral series and that the minimal just-noticeable difference in wave-length is one millimicron. This latter value is realized at three points in the spectrum: 588.5 $m\mu$, 507.0 $m\mu$, and 489.4 $m\mu$, respectively. The rate of variation of chroma with respect to wave-length shows a maximum at these points, but this rate becomes very low at the extremities of the spectrum, particularly at the long-wave (red) end. At the present time there are no available data from which we can compute a scale of pure hue, eliminating saturation effects, and such a scale could not be established in terms of homogeneous stimuli alone.

Saturation as a Function of Wave-Length. Direct introspection shows that the saturations of the spectral colors are lower in the middle than at the ends of the spectrum. Jones and Lowry (32) have determined the saturations of the spectral colors by stepping off just-noticeable differences in saturation from a gray of the same brilliance along a number of constant hue lines. As shown in Table 3, the spectral saturation shows a minimum at about 575 $m\mu$, having a value of 16 at this point, as compared with 23 at the extremities. Certain facts of color mixture and phenomena of chromatic flicker confirm these conclusions. In order to

TABLE 2

Spectral Chroma Scale

No.	λ(mμ)	dλ	No.	λ	dλ	No.	λ	dλ
1	700.0	....	44	576.5	1.4	87	490.4	1.1
2	678.0	22.0	45	75.2	1.3	88	89.4	1.0
3	65.0	13.0	46	73.7	1.5	89	88.2	1.2
4	59.0	6.0	47	71.7	2.0	90	87.0	1.2
5	54.0	5.0	48	70.1	1.8	91	85.8	1.2
6	49.5	4.5	49	68.4	1.7	92	84.5	1.3
7	46.0	3.5	50	66.6	1.8	93	83.2	1.3
8	42.8	3.2	51	64.8	1.8	94	81.7	1.5
9	40.2	2.6	52	63.0	1.8	95	80.0	1.7
10	37.8	2.4	53	61.1	1.9	96	78.2	1.8
11	35.5	2.3	54	58.6	2.5	97	76.5	1.7
12	33.1	2.4	55	57.0	2.6	98	75.0	1.5
13	30.0	3.1	56	54.4	2.6	99	72.9	2.1
14	26.5	3.5	57	51.8	1.6	100	70.5	2.4
15	23.0	3.5	58	49.1	2.7	101	68.2	2.3
16	20.0	3.0	59	46.1	3.0	102	65.8	2.4
17	17.3	2.7	60	43.0	3.1	103	63.6	2.2
18	14.9	2.4	61	39.8	3.2	104	61.2	2.4
19	12.5	2.4	62	36.5	3.4	105	58.7	2.5
20	10.2	2.3	63	33.2	3.3	106	56.5	2.2
21	08.0	2.2	64	30.1	3.1	107	54.4	2.3
22	06.0	2.0	65	27.1	3.0	108	52.1	2.3
23	04.1	1.9	66	24.2	2.9	109	50.0	2.1
24	02.3	1.8	67	21.4	2.8	110	48.0	2.0
25	600.6	1.6	68	19.1	2.3	111	46.0	2.0
26	599.0	1.6	69	16.8	2.3	112	44.2	1.8
27	97.4	1.6	70	14.6	2.2	113	42.5	1.7
28	95.9	1.5	71	12.6	2.0	114	40.8	1.7
29	94.5	1.4	72	10.6	2.0	115	39.0	1.8
30	93.1	1.4	73	08.0	1.6	116	37.2	1.8
31	91.8	1.3	74	07.0	1.0	117	35.3	1.9
32	90.5	1.3	75	05.4	1.6	118	33.3	2.0
33	89.5	1.0	76	04.0	1.4	119	31.3	2.0
34	88.5	1.0	77	02.6	1.4	120	29.3	2.1
35	87.5	1.0	78	01.3	1.3	121	27.0	2.2
36	86.4	1.1	79	500.0	1.3	122	24.8	3.2
37	85.3	1.1	80	498.7	1.3	123	22.3	2.3
38	84.0	1.3	81	97.4	1.3	124	19.7	2.8
39	82.7	1.3	82	96.1	1.3	125	16.7	3.0
40	81.5	1.2	83	94.8	1.3	126	13.8	2.9
41	80.3	1.2	84	93.7	1.1	127	10.4	3.4
42	79.1	1.2	85	92.6	1.1	128	405.8	4.6
43	77.9	1.2	86	91.5	1.1	....	....	....

generate a scale of equally saturated colors, it would be necessary to add "white"—or else a certain proportion of complementary stimuli—to all wave-lengths, with the exception of 575mμ.

TABLE 3

SATURATION SCALE DATA

Wave-length, mμ	440	470	490	540	575	605	640	680
No. of steps	23	22	19	19	16	20	23	23

Chroma as a Function of Wave-Length Composition. When heterogeneous stimuli are employed, the manner of determination of chroma becomes highly complex. There are, of course, an infinite number of possible combinations of wave-lengths, showing either a continuous or a discontinuous distribution of intensities with respect to the wave-length scale. The principles which must apply to *continuous* distributions can be developed from those for a finite number of discrete wave-lengths. The simplest classification is, of course, that involving mixtures of only two homogeneous stimuli.

Two-component mixtures. If we select a stimulus of single fixed wave-length at the long-wave end of the spectrum and mix with it a second stimulus having any desired other wave-length, changing the latter gradually towards shorter values, the following phenomena are observed. The hue which is evoked by the mixture varies continuously from red through orange, yellow, yellow-green, and green until the variable stimulus reaches a wave-length of approximately 492 mμ (corresponding to a blue-green). At this point, if the intensities of the two stimuli are properly chosen, the subjective result is a gray—or some achromatic color—and the stimuli in question are said to be *complementary*. Up to a wave-length of about 580 mμ, the saturation of the mixture is practically identical with that belonging to the spectrally evoked color of the same hue. However, with shorter wave-lengths than 580 mμ, the saturation is less than spectral. The exact hue and saturation depend, in any case, upon the ratio of the two stimulus intensities, as well as upon their respective wave-lengths. For any pair of wave-lengths, the hue can be made to vary continuously from that belonging to one to that normal to the other, by passing through the whole gamut of possible intensity ratios. When the wave-lengths are complementary, variations in intensity ratio can yield only two hues, namely, those belonging to the given component wave-lengths, respectively. That hue will appear, the stimulus for which is in excess of the complementary ratio. The saturation will increase as the intensity relations depart from the ratio in question.

If the variable wave-length in the two-component mixture is carried *beyond* the complementary point, the hue attaching to the mixture will correspond to that for a wave-length which is intermediate between the variable stimulus and the *short-wave* end of the spectrum, or else it will belong to the series of purples (*vide infra*). If, in this experiment, the long-wave (fixed) stimulus be made progressively shorter, the corresponding complementary point will also shift towards shorter wave-lengths. When the long-wave stimulus reaches about 567 mμ, the complementary will be located approximately at the extreme short-wave end of the spectrum. The experiment can, of course, be repeated with the "fixed" wave-length at the short-wave end of the spectrum and in this case the results will follow principles that are logically identical with those above stated, but involving, primarily, the range of hues between the violet and its complementary, the yellowish green.

It appears, from the above-stated facts, that, with respect to two-component mixtures, the spectrum is divisible into three ranges of wave-lengths. Members of one extreme range find their complementaries within the other extreme range, but the middle range of wave-lengths has no complementaries within the spectrum. Table 4 shows the pairs of complementary wave-lengths as found by Sinden (61). According to Priest (51), complementary wave-lengths are related to one another in harmony with the equation for a rectangular hyperbola. The intensity ratios, which are required, at complementary wave-lengths, to yield an achromatic color, vary with the identity of the complementary pairs. If these intensities are measured photometrically, it appears that the short-wave stimuli have, in general, a much higher chromatic power than the longer-wave stimuli. Thus, the photometric amount of violet which must be added to the yellow-green to yield a gray is approximately one per cent of the brightness of the yellow-green.

The extra-spectral hues or "purples." As has already been noted, mixtures of long- and short-wave stimuli, preferably taken from the spectral extremities, yield a range of chromas which are ordinarily called purples, although they include the psychologically primary red and a few slightly yellowish reds. Since the extreme spectral reddish hue is somewhat yellowish, it is necessary to add a small proportion of short-wave stimulus to secure the psychological red. According to the work of F. O. Smith (62), there are 28 just-noticeably different chromas in this extra-spectral series, making a total of 156 maximally saturated colors around the entire hue cycle.

Three-component mixtures. Experiment shows that the extra-spectral series of two-component mixtures has complementaries within the mid-region of the spectrum, which finds no homogeneous complementary stimuli; the purplish hues are, in general, complementary to the greenish ones. A complementary combination of this sort obviously comprises a *three-component mixture.* Greenish or purplish colors of many different saturations can be evoked by appropriate choices of relative intensities between the mid-spectral and the extreme-spectral pair. From these and previous considerations, it can be seen that such a combination of three stimuli can be varied, in intensity ratios only, so as to match all possible hues and a wide range of saturations. The intensity of one or even two members of the triad can, of course, be made zero, yielding a two-component or a homogeneous stimulus, respectively. This three-component system, consisting of a mid-spectral and two opposite extreme-spectral stimuli, therefore furnishes us with *a method for formulating the entire system of colors as a function of three independent variables having fixed wave-length characteristics.*

This system is capable of dealing with the achromatic as well as with the chromatic features of color but, if the former are excluded, the facts can be stated in terms of two independent variables, expressing the *ratios* rather than the absolute values of the three intensity components. Such a two-variable scheme is the basis of the ordinary color-mixture triangle. We can represent all possible combinations of three variables by means of a three-dimensional system of Cartesian coordinates, as shown in Figure 6. Any point in this system will have three parameters, corresponding to the respective intensity measures of the three mixture-components. If desired, these intensities can be expressed either in radiometric or in photometric units, but, in practice, a different basis of measurement is ordinarily used in order to avoid the asymmetry which results from the very unequal chromatic powers of the several primary wave-lengths. The preferred scheme assumes that the proportions of the three stimuli which combine to yield a gray are metrically equal. This means that a line forming a central axis between the coordinate planes of Figure 6 represents the locus of all achromatic colors or of the corresponding mixtures. Now, if we wish to avoid absolute measures, we can confine our attention to the angular projections of points in the total system upon a surface *RGB* which is perpendicular to the achromatic axis. This projection surface is evidently bounded by an equilateral triangle similar to the conventional color-mixture diagram (see Figure 7).

TABLE 4

WAVE-LENGTHS OF SPECTRAL COMPLEMENTARIES

(According to Sinden)

Given wave-length	Complementary wave-length	Nine observers
Extreme red	496.5	497.1
609	493.5	
609.5		494.1
591	490.	
586	487.5	
581		485.5
580	482.5	
578.5	480.5	
576.5	477.5	
575.5	474.5	
574	472.	
573	466.5	
572	459.	
570.4		
570	Extreme violet	

In order to locate specific colors with respect to a system of this sort, we must establish points or lines within it, corresponding to colors which are determined by some criterion other than that of mixture proportions. The most important of such independently specified systems of colors is undoubtedly that provided by the spectral series. The spectrum can be "calibrated" in terms of a three-component system by matching each of the spectral hues with an appropriate proportion of intensities in the three-component mixture. In practice, it is found impossible to secure complete *saturation* matches in all cases. However, perfect matches can be obtained if an appropriate amount of complementary is added to those spectral stimuli which fail otherwise to match in saturation. The amount of such admixed complementary enables us to compute a theoretical locus for the corresponding spectral color, *outside* of the area bounded by the triangle. Such a locus is significant of negative values in the original Cartesian system of coordinates. The necessity for such a placement of the majority of the spectral colors eventually leads us to recompute the three-component scheme in terms of primaries which have super-spectral saturations. This reconstruction permits us to assign a locus to the spectral series, as a whole, within the triangle, as shown in

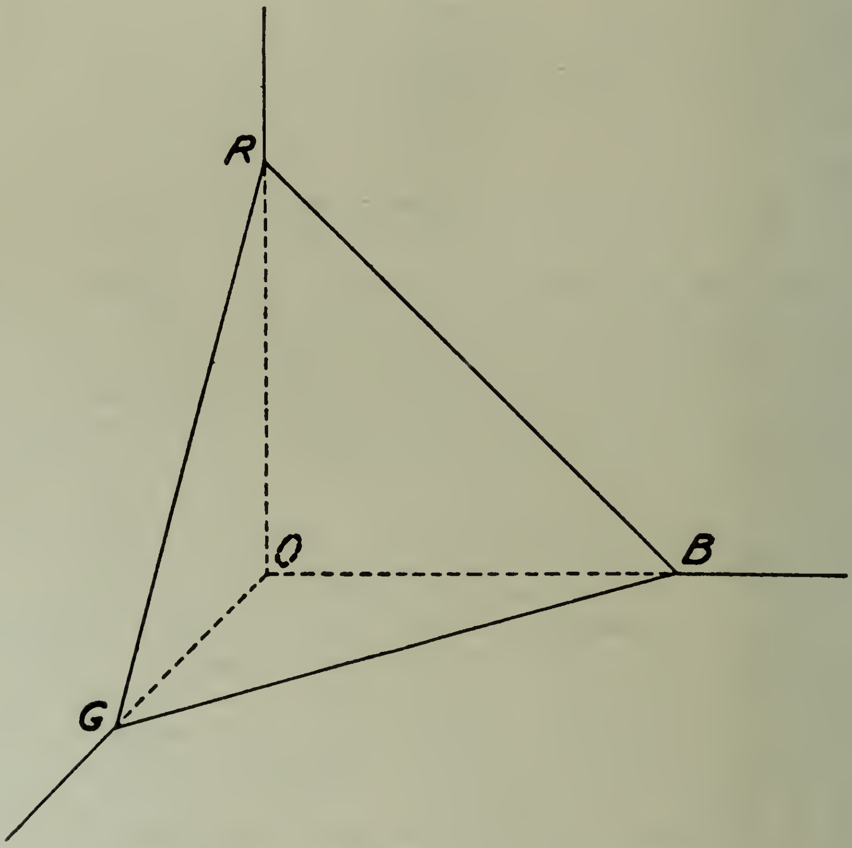


FIGURE 6

THE COLOR-MIXTURE TRIANGLE, IN RELATION TO A TRIDIMENSIONAL COORDINATE SYSTEM

(To be used in: Troland, L. T. The principles of psychophysiology. New York: Van Nostrand.)

Figure 7. Other systems of colors, such as those produced by black-body radiation at various temperatures, can be plotted in a similar way.

Each point in the plane of the color-mixture triangle represents a specific chroma, and the consequences of mixing the stimuli corresponding to two or more chromas can be determined by purely mathematical considerations applying to the triangle. If stimuli corresponding to two separate points in the triangle are mixed in chromatic intensities having the ratio $x:y$, then the chroma resulting from such mixture will be represented at the *centroid* of the two-point system. This centroid has the position of an imaginary center of gravity, as if x and y were specifically localized masses. Mixtures involving more than two components can be handled by successive combinations in pairs, in accordance with the same principle. In this way, the color-mixture triangle provides a method for studying the consequences of any kind of stimulus mixture, no matter how complex.

Mixtures consisting of a continuous distribution of intensities along the wave-length scale can be handled by dividing such continua into a finite number of regions, each of which does not appreciably exceed in span the just-noticeable difference in wave-length in its own particular position. These artificial components can then be treated by the principle of the color-mixture triangle. Moreover, Priest (50) has

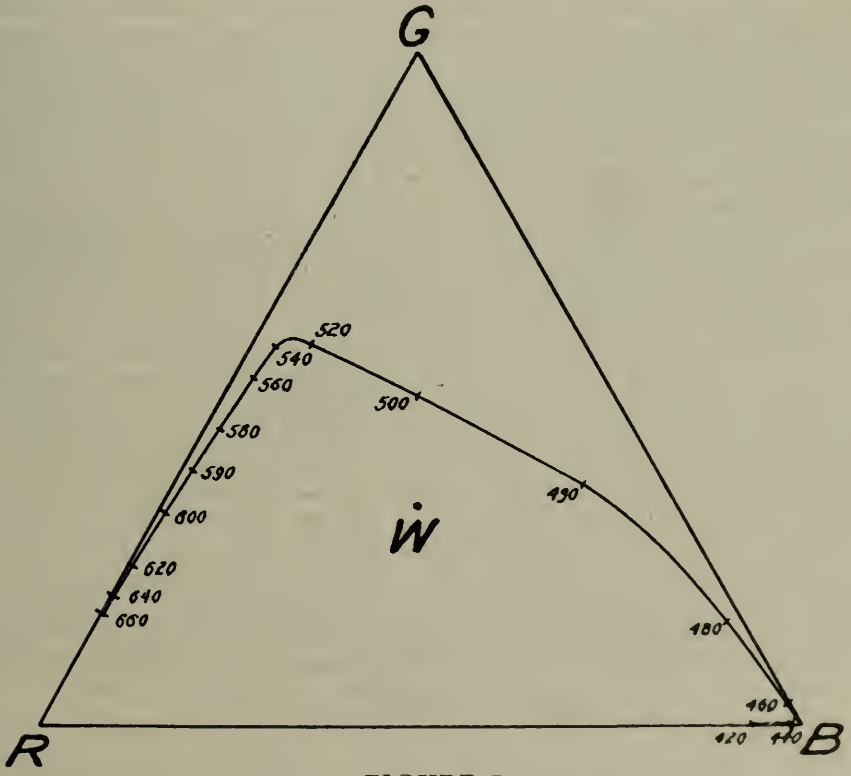


FIGURE 7

The Color-Mixture Triangle, With the Locus of the Spectral Colors

(After Ives. To be used in: Troland, L. T. The principles of psychophysiology. New York: Van Nostrand.)

shown that if the centroid of the intensity-distribution curve, plotted in photometric terms, is determined, it lies at a wave-length whose hue very closely approximates that belonging to the mixture. The saturation will, of course, be lower, the greater the spread of the stimulus in the spectrum.

The relative constitution of any chroma which is represented in the triangle in terms of the three selected mixture-primaries is given by the lengths of the three perpendiculars dropped from the point representing the chroma in question to the three respective sides of the triangle. A study of the geometry of the triangle shows that the linear sum of these three perpendiculars is a constant, regardless of the position of the given point in the triangle. The results obtained by calibrating the spectrum in terms of a three-component mixture can be plotted in the form of three so-called chromatic excitation curves shown in Figure 8, the corresponding numerical values being given in Table 5. These curves reveal the proportions of the three components which are required for a color-match at each wave-length, the units being so chosen as to make the three areas equal, in harmony with the requirement that equal amounts of the three primaries should yield a gray or white. The corresponding numerical values are given in Table 4.

It should be noted that there is a wide range of freedom in the choice

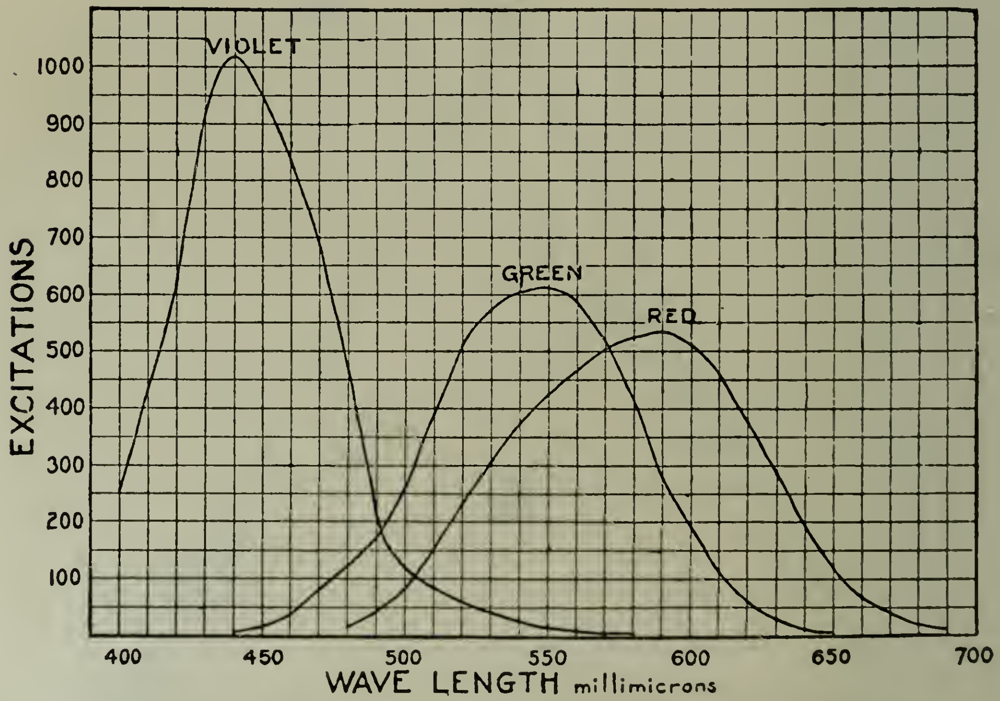


FIGURE 8

The Elementary Color Excitations as a Function of Wave-Length

(According to Weaver. From: Report of the Colorimetry Committee of the Optical Society of America, L. T. Troland, Chairman. *J. Opt. Soc. Amer.*, 1922, **6.**)

of the three primaries for the general type of system which we have just been discussing. However, the best selection seems to be an extreme long-wave stimulus (spectral red), a wave-length corresponding to the spectral green at about 505 mμ, and either an extreme short-wave stimulus or one evoking a blue, at 475 mμ. The choice of a "blue" in place of a "violet" has an advantage on account of the psychological primacy of the former, but requires a recrudescence of the red primary in the violet to account for the reddish hue of the latter. It is also possible to recompute the system in terms of an extra-spectral stimulus corresponding to the psychologically primary red. Any set of primaries thus determined can be designated as *stimulus*, in opposition to *psychological* primaries.

Chroma as a Function of Intensity. The color-mixture system, which we have just considered, really establishes no fixed relationship between hue and saturation (as subjectively measured) and the stimulus variables. The scheme of three-color mixture-matching resembles the visibility procedure in that subjective matches are made the criteria for determining an equivalency between physically different stimuli. The color triangle takes no cognizance of absolute intensities, but the latter have an important bearing upon absolute hue and saturation. Starting at the lowest or absolute threshold intensities, increase in intensity is accompanied by the following progression of hue and saturation changes. Using homogeneous stimuli, we find, in the minimal intensity range, a tendency towards a very unsaturated or even a wholly achromatic color. This tendency is more marked in the case of the short than in that of the long waves. Red appears in practically undegraded saturation at the lowest intensities, whereas the normal wave-lengths for blue evoke a pure gray under such conditions. The intensity range over which this gray appears is commonly known as the *photochromatic interval* (46). The magnitude of this interval is less in the center than in the periphery of the visual field and increases markedly with scotopic dark-adaptation in the periphery. The intensity range of the photochromatic interval corresponds approximately with the degree of dominance of the scotopic visibility curve, which is associated with a purely achromatic visual effect.

As the intensity is increased, the saturation of the originally desaturated color at first increases but begins to diminish again at a higher intensity level. It is commonly stated that high intensities can be found at which homogeneous stimuli yield purely achromatic colors or "white," but this assertion does not appear to have been tested accurately. In addition to the saturation decrease, with increasing intensity in the higher intensity range, there is a systematic shift of the hues in relation to wave-length, an effect which is known as the Bezold-Brücke phenomenon. This can be demonstrated through a simultaneous comparison of the hues evoked by stimuli of the same wave-lengths but of different intensities. The colors, on either side of the psychological red and green, shift in hue towards the psychological blue or yellow. In this process, three "invariable" points are demonstrable within the spectrum. According to results obtained by

TABLE 5

SPECTRAL COLORS IN TERMS OF ELEMENTARY EXCITATIONS

These values are for an equal energy spectrum. The relative magnitudes of the three elementary excitations have been chosen so that the curves for average noon sunlight have equal areas; that is, if the percentage values are plotted on a trilinear diagram, sunlight falls in the center. The absolute excitation values are based upon a convenient arbitrary unit. The percentage values are given for the red and violet only, since those for the green can be found by subtracting the sum of the other two values from 100 in each case.

Wave-length mμ	Excitations			Percentages		Wave-length	Excitations			Percentages	
	Red	Green	Violet	Red	Violet		Red	Green	Violet	Red	Violet
400	...	...	253	...	100	550	424	612	18	40.2	1.7
410	...	...	433	...	100	560	466	578	11	44.2	1.0
420	...	...	614	...	100	570	505	517	7	49.5	.7
430	...	...	915	...	100	580	520	415	4	55.4	.4
440	...	7	1019	...	99.3	590	535	296	..	64.3	..
450	...	16	950	...	98.3	600	510	196	..	72.2	..
460	...	38	842	...	95.7	610	462	113	..	80.4	..
470	...	81	697	...	89.6	620	375	59	..	86.3	..
480	14	122	473	2.3	77.6	630	285	29	..	90.8	..
490	41	169	220	9.5	51.1	640	195	10	..	95.1	..
500	83	260	123	17.8	26.4	650	118	3	..	97.5	..
510	151	391	87	24.0	13.8	660	68	...	..	100	..
520	233	510	61	29.0	7.6	670	40	...	..	100	..
530	307	572	43	33.3	4.7	680	22	...	..	100	..
540	373	603	29	37.1	2.9	690–750	27	...	..	100	..

D. McL. Purdy,[2] working in my laboratory, these points lie about 575 mμ, 507 mμ, and 474 mμ, respectively, for an intensity change between 100 and 1000 photons. It is to be presumed that there is a third invariable in the extra-spectral series corresponding with the psychologically primary red. The manner of variation of apparent wave-length between 100 and 1000 photons over the whole spectrum is shown in the accompanying curve (see Figure 9) taken from Purdy's work. In this system, the yellow and blue acts like points of stable equilibrium, whereas the red and green behave as points of unstable equilibrium, the hues shrinking away from the latter towards the former regions.

Spatial Determinants of Chroma. The chromatic attributes vary as functions of the angular size, shape, and position of the stimulus within the field of vision. They are also radically influenced by the state of stimulation in outlying retinal regions. Hue and saturation are markedly affected by reductions in the size of the stimulus. With very small angular sizes, certain common wave-length discriminations, such as that between "blue" and "green," are lost entirely, particularly with a brilliant surrounding field. Purely chromatic patterns disappear much more

---

[2] Purdy's work is as yet unpublished.

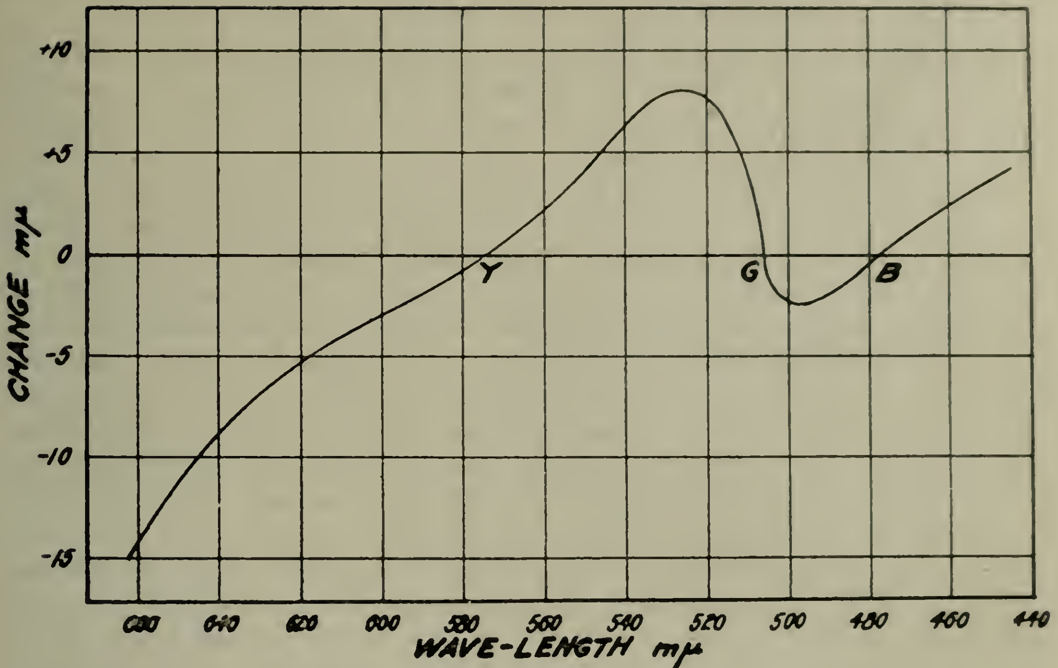


FIGURE 9

Graph of the Bezold-Brücke Effect

(According to Purdy. To be used in: Troland, L. T. The principles of psychophysiology. New York: Van Nostrand.)

readily than do achromatic ones.[3] These phenomena appear not to have been studied quantitatively.

Chromatic zones. The changes of hue and saturation with respect to position in the visual field follow a law which is formally similar to that governing the Bezold-Brücke phenomenon. In this relation the center of the field corresponds to medium, and the periphery to extremely high intensities. This means that if a stimulus, either homogeneous or heterogeneous, is moved, centrifugally, it must lose saturation and its hue must shift in the manner which we have considered. Certain red, green, yellow, and blue colors remain invariable, while intermediates move towards the blue or the yellow. In general, the rate of desaturation of the red and green, with respect to angular departure from the center, is greater than that of yellow and blue. According to the work of Baird (4) and others, if the stimuli for the members of these respective color pairs are of the composition required for complementary combination, they lose their chromas at the same peripheral position. These facts have led to the notion of *color zones* within the visual field, according to which the center of vision has full trichromatic response; an intermediate more peripheral zone has yellow-blue vision only, while the extreme periphery yields achromatic colors alone. However, the studies of Ferree and Rand (18) indicate that these zones, if they exist at all anatomically, are not rigidly fixed but are dependent upon the intensity as well as upon the purity of the stimulus. They find that all hues, with the exception of green, are evocable under the most favorable conditions, even at the extreme periphery.

---

[3] Dr. Theodore Karwoski is now working on the problem of pure chromatic acuity in my laboratory.

Chromatic contrast. The chroma which is evoked by a given stimulus acting upon a restricted test field is influenced markedly by the state of excitation upon adjacent areas. The general principles which are involved can be formulated most concisely in terms of subjective variables, a procedure which is legitimate because contrast effects are evidently conditioned not directly by the action of the outlying stimulus but indirectly by their excitation consequences. Broadly stated, chromatic contrast displaces the contrasting colors in the direction of a *complementary relationship.* This principle can be interpreted as implying saturation as well as hue changes. Thus, if the test field contains an unsaturated red while the surrounding field is a saturated red, the test field will become less saturated than in the absence of the latter, through contrast action, as if a certain quantity of surrounding field complementary had been added to it. On the other hand, if the test field is normally achromatic, the reddish environment will induce a bluish-green tinge, while, if it were initially blue-green, its saturation would be increased through the contrast action.

It has been demonstrated by Kirschmann that chromatic contrast is at a maximum when the test and surrounding fields are of equal brilliance ("Kirschmann's law"). Chromatic contrast is favored by similarity or uniformity of texture between the two fields and by an absence of dark boundary lines.

Chroma as a Function of Time. No adequate studies appear to have been made upon the laws of rise or decay of the chromatic attributes for given attained values of the brilliance. However, there is plenty of evidence that the chromatic constituents of "white" grow and decay at different rates. The front edge of a moving white object is reddish, while the rear edge is bluish. Broca and Sulzer (7) found the rate of rise of spectral stimuli to increase in the order: green, red, blue; but Bills (5), using photometrically equal intensities, noticed the highest speed at 580 mμ and the lowest at 563 mμ. Ives (28) found that the inertia decreases progressively with diminishing wave-length, so that mixed red and blue or red and green can be separated spatially by means of an oscillatory movement of the stimulus over the retina. However, in these effects, we are apparently dealing not with the chromatic attributes alone but with the total color.

Chromatic flicker can be evoked by alternating, upon the same retinal area, two stimuli which differ in wave-length composition but which are photometrically equal. This is the principle of the well-known *flicker photometer,* when a brilliance match has been attained. The critical flicker-frequency is at a minimum when the two stimuli have the same photometric value, and the magnitude of this minimum increases with the chromatic difference between the two colors. When each of the spectral colors is alternated with an achromatic color of about the same brilliance, a minimum is found at 575 mμ (70), corresponding with the minimum of saturation at this point. Ives (29) has shown that the quantitative laws which govern these phenomena are in harmony with those to be expected from the specific inertias or diffusivities which attach to the different spectral stimuli.

When chromatic fusion is attained, the resulting chroma is the same as would be produced by a temporally smooth distribution of the same stimulus combination. This is the principle of chroma mixture with rotating sector disks, and corresponds to the Talbot-Plateau law for brilliance.

Chromatic adaptation. Continued exposure of a fixed retinal region to a chromatic stimulus is accompanied by a loss of saturation which, in certain cases, may be complete. It has been claimed by Voeste (73) and others that there is also a shift in hue similar to that observed in the Bezold-Brücke effect. However, C. H. Langford and the writer have shown that there is actually no hue change (72), the effect observed by Voeste being the Bezold-Brücke phenomenon itself. My experiments also demonstrate that, when the initial saturation is high, even equilibrium adaptation does not completely destroy the chroma (as claimed by Hering). The asymptotic saturation is uniformly greater, the higher the original saturation, although if the latter is very low the chromatic effect may be carried below the threshold.

Chromatic after-images result when adaptation, as above described, occurs on a restricted visual area. If the original stimulus is withdrawn and the visual field is observed in the dark, the after-image will ordinarily be complementary in hue. This is also the case when we employ a reacting stimulus that would normally yield an achromatic color. However, if the original stimulus is left in place but is lowered in intensity, a wide variety of chromatic changes can be induced. In this so-called *dimming effect,* which Dr. Karwoski and I (71) have studied very extensively, there is a great predominance of purplish hues. The change in hue is greater, the greater the fractional dimming and the higher the intensity of the original stimulus or the greater the extent of the adaptation. The predominance of purples, together with the relations which are found between the chromatic effects and the original hues, makes it impossible to explain the *chromatic dimming phenomena* in terms of the usual complementary after-image principle. These phenomena seem to rest upon differences in the rates of change of separable chromatic components subsequently to the dimming. Apparently, the green component falls more rapidly than do the red and the blue, at least under conditions of selective chromatic adaptation.

If such transitional phenomena are neglected, it appears that chromatic adaptation can be expressed in terms of a "coefficient law" applying to separable chromatic constituents. Any case of chromatic adaptation involves an alteration in the proportions of these constituents, and such an alternation combines with any stimulus so as to change the chroma which it evokes, as if the proportions of the color-excitation curves of Figure 8 had been semi-permanently modified.

Color-Blindness. Important differences are found between individuals in regard to the psychophysics of both the chromatic and the achromatic attributes. In addition to relatively small fluctuations, such as occur in any biometric data, there are larger variations which indicate distinct types or species, different from the normal. The most striking of such aberrant types are those corresponding to what is commonly called color-blindness. Color-blindness, in its most extreme manifestations, is at most only *chroma blindness* and in the commonest forms is only red-green blindness. Although, according to Schjelderup (59), there are at least 18 significant species of color-blindness, the most important types are: (1) anomalous trichromatism; (2) dichromatism, including (*a*) protanopia, (*b*) deuteranopia, and (*c*) tritanopia; and (3) monochromatism (46, Part 2).

The *anomalous trichromats* experience a range of chromas which is qualitatively similar to that given to the normal trichromat, and they require three separate stimuli to match the spectral hues by mixture. However, their mixture ratios depart rather radically from the normal. Thus, in matching a spectral yellow with a mixture of long- and intermediate-wave stimuli (red and green), either an excess of the former or of the latter may be required, indicative of the protanomalous and deuteranomalous types, respectively. The protanomalous suffers a reduction in the brilliance (visibility) of the red as well as in its color-mixture power. The anomalous condition is characterized by generally high thresholds, easy visual fatigability, increased contrast effects, and other evidences of functional unbalance.

In the first two forms of *dichromatism,* chromatic experiences are reduced to terms of the two psychological primaries, yellow and blue. About four per cent of males and about one-tenth of this proportion of females suffer from this defect, in the human species, and the condition exhibits the properties of a Mendelian sex-linked character. Tritanopia is a much rarer condition in which the yellow and blue psychological primaries are absent. It is apparently attributable to retinal disease rather than to hereditary variables. Dichromats of any type can color-match the entire spectrum by means of only two mixture-components, so that their color systems can be represented by lines, in place of the usual plane figure of the color-mixture triangle. The dichromats attach an achromatic color to a certain homogeneous stimulus, lying at about 492 mμ for the protanopes, at about 498 mμ for the deuteranopes, and at about 583 mμ for the tritanopes.

The principal difference between the protanopic and deuteranopic types appears to consist in the radical reduction in visibility for the former at the long-wave end of the spectrum, in comparison either with the normal trichromatic or with the deuteranopic types. This means also that in color-mixture equations, matching intermediate spectral colors, the protanopes require more long-wave intensity than normal. It is an important fact that color-matches which are established by normal trichromats are valid for any dichromat, which makes it possible to regard dichromatism as derived from trichromatism by the simple loss of a unit chromatic process. Trichromatic excitation curves have been computed by König (35) and others in such a manner that the "red" curve is identical with the long-wave curve of the deuteranope, and the "green" curve is the same as the long-wave curve of the protanope. However, cases of monocular dichromatism indicate that the chromatic experiences are qualitatively alike for both types, so that they fail to correspond subjectively with the old fashioned distinction between "red"- and "green"-blindness.

In monochromatism, color experience is limited to a purely achromatic scale of grays. The visibility curve is of a scotopic type and is ordinarily accompanied by a condition of day-blindness or photophobia. These facts indicate the functional absence of the photopic system in its entirety, corresponding to an inoperative state of the retinal cones. The reverse of this condition, involving a loss of scotopic or rod function, is well known under the name of hemeralopia or *night-blindness.*

VI. Binocular Relations of Color

Under the commonest conditions of vision, the differences between the stimulus patterns which are presented to the two eyes, respectively, have relatively little bearing upon color phenomena, although they profoundly affect the experience of depth. However, when regularly reflecting surfaces are involved or under artificial conditions, large corresponding areas of the two retinas can be stimulated quite differently. In such cases, special subjective phenomena of color are observable. The most familiar of these effects is that known as *luster,* the peculiar glint which characterizes metallic and other polished surfaces. It is attributable to the difference in intensity between corresponding areas of the two retinal images.

Rivalry and Fusion. When, under artificial circumstances in the laboratory, large retinal areas are stimulated discrepantly in this way, experience is apt to show an alternation between the colors, textures, or patterns belonging monocularly to the respective stimuli. This effect is called *retinal rivalry.* The large number of experimental studies which have been made upon it show that its liability to occur and the rate of alternation increase with increasing difference between the discrepant stimuli (56). Rivalry may take place within restricted portions of the visual field, as well as sometimes involving the total field.

In the absence of rivalry, the so-called *fusion quality* in consciousness usually corresponds to a mean between the colors belonging to the two monocular stimuli. In the case of discrepant brilliances, the "fused" color is naturally less brilliant than the higher of the two monocular brilliances. This effect is so opposed to the normal expectation, viz., that the two intensities should add arithmetically, that it has been called Fechner's *paradox,* after the man who first studied it quantitatively.

Binocular Color Mixture. When the two discrepant stimuli are such as to yield chromatically different colors, a similar principle applies, the color tending to lie midway between those normal to the individual members of the given stimulus pair. The majority of observers have found it difficult to secure satisfactory fusion when the chromatic differences were large. However, the recent work of Hecht (20) shows that, if the stimulus pattern is favorably chosen, a resultant color can be obtained even with the greatest possible chromatic discrepancy. Under these conditions, complementaries yield "white," while red and green evoke yellow (see next chapter).

Binocular Flicker. When stimuli are alternated between the two eyes, so that the dark intervals for one eye are filled by a positive stimulus at the other eye, the critical flicker-frequency is only slightly different from what it would be if only a single eye were involved. This seems to mean that, at ordinary critical flicker-frequencies, the locus of the fusion process is retinal rather than cerebral. The laws which govern binocular flicker have been studied in detail by Sherrington (60).

VII. Monocular Laws Governing Visual Form

We have noted that the subjective visual field has a three-dimensional structure but, that it is nevertheless capable of being viewed as projected upon a plane which is perpendicular to the empirical line of sight. The pattern of colors appearing upon this plane at any moment is primarily determined by the distribution of stimuli or of stimulus characteristics over the composite retina (right and left eyes together). If one eye is closed so that we are dealing with actual monocular vision, we can avoid the complications which are due to binocular disparation. In the field before the eyes, considered objectively, the corresponding pattern of objects or of stimuli must be expressed in angular terms with respect to the average position of the eyes, or else the configuration of the object field must be projected, towards this point, upon a surface which is preferably spherical and axial with the average line of regard.

Visual Acuity. The relation of the patterns, which appear in the visual field, to the object field outside of the eyes is characterized by a finite resolving power or definition, analogous to that of a lens system. This means that patterns of angular detail smaller than a limiting value receive no specific representations in the corresponding visual consciousness. This finite resolving power of the visual system is usually measured in terms of an ability to discriminate between two parallel dark lines, separated by a white interspace. The *acuity* measure is taken as the reciprocal of the angular separation of these lines, measured in minutes of arc, and the standard of good acuity at favorable illumination levels is taken as unity on this scale.

Visual acuity varies as a function of practically every stimulus attribute. Thus, it alters with wave-length and with wave-length composition, being greater for mid-spectral stimuli than for the extreme long or short waves of the spectrum. These variations can be attributed in part to the chromatic aberration of the eye, which causes rays of different wave-lengths to come to a focus differently upon the retina. Optical phenomena of diffraction and scattering of light, which are involved in the formation of the retinal image, also vary with wave-length and affect the sharpness of vision. Acuity is particularly bad with mixtures of wave-lengths from the two spectral extremities, as in the case of the rays which are transmitted by a purple cobalt glass. This is due to the impossibility of bringing both of these classes of rays to a sharp retinal focus at the same time.

Acuity also varies in a very important way as a function of the stimulus intensity. When the retinal illumination is decreased below a level of about 60 photons (cf. next chapter), the acuity also begins to decrease, becoming extremely poor near the absolute threshold. The acuity is constant over a considerable range of higher intensities in the region where the relative intensity discrimination threshold is approximately constant, but diminishes at still higher intensities, which yield "glare." Measurements show that, in the lower intensity range, the value of the acuity is a logarithmic function of the intensity value, according to a law having two sections with different constants. The details and explanation of these acuity functions, in relation to the retinal process, are presented in the ensuing chapter by Dr. Hecht. It should be noted that, although the basis of these variations is primarily retinal, acuity is favored at higher intensity levels by the constriction of the pupil, a small lens diaphragm being helpful to good definition in optical systems.

Acuity also varies radically with change in the position of the stimulus pattern within the visual field, falling off rapidly from the center towards the periphery (76). This is to be attributed, in part, to the inferior optical resolving power of the eye in the periphery, but even more to the greater minuteness of retinal structure in the central region, together with the superior detail of the corresponding nerve connections.

Distortions of Form. The visual system shows characteristic distortions of form which are similar to those manifested by uncorrected lens systems. Among these are included barrel distortion, coma, and astigmatism. Other distortions reveal peculiarities of structure in the ocular tissues. Thus, the "rays" seen on stars and similar brilliant points of light are attributable to directional scattering of the stimulus by the radial laminations of the crystalline lens. Halos are sometimes observable around bright objects, in consequence of diffraction by the cells of the cornea. With short-wave rays, sharply localized within the pupil, a snake-like figure, corresponding to the retinal blood-vessels, can be perceived. Numerous other *entoptic phenomena* have been described by Purkinje (52) and his successors.

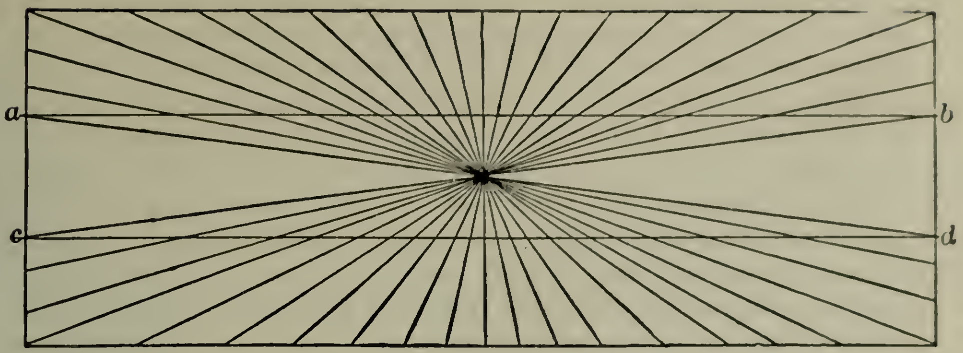
One of the most interesting of these consists in the so-called "blue arc phenomenon." This consists of curved shafts of luminescence which can be seen shooting off from a patch of brilliant color (preferably red) presented against a dark outlying field. These arcs connect the primary stimulus patch with a position in the visual field corresponding to the exit of the optic nerve-fibers and seem to be due to nerve-currents passing across the retinal surface from the primarily stimulated receptors. Gertz (19) and others have regarded the mechanism of the secondary stimulation as electrical, but Mrs. Ladd-Franklin (37) argues that the nerve-currents must actually generate radiant energy, since the blue arcs have a definite after-image. The blind-spot itself presents another case of failure of the subjective visual pattern to correspond with the configuration of the object field, what is presented in the blind-spot being determined mainly by the stimuli acting upon the retina outside of the papilla (exit place of the optic nerve).

Geometrical optical illusions. Another species of visual distortion of form is comprised by the so-called geometrical optical illusions. These consist of figures made up of various combinations of lines or areas, in which the apparent relations of the given units differ from those which can be proven to exist between the corresponding stimulus or object elements. Thus, in the familiar Müller-Lyer illusion, two objectively equal distances are represented subjectively as unequal because of differences in the directions of arrow-heads which are attached to the ends of the two distances or lines. Straight lines may appear curved or broken, parallel lines may be represented as non-parallel, and so on. Such discrepancies between the objective and the subjective patterns appear to arise from peculiarities in the relationships of the various constituents of the given configurations. Thus converging arrow-points have a shortening effect, while diverging ones act in the opposite manner. A number of these patterns are shown in Figure 10. The causes of these distortions, in general, are to be sought in cortical rather than in intra-ocular functions.

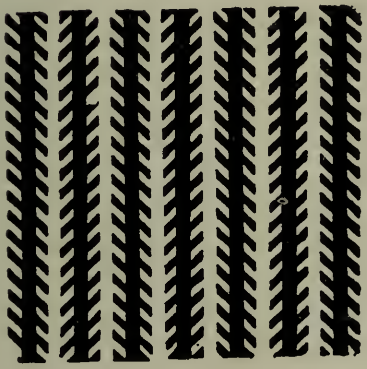
VIII. Depth and the Binocular Determination of Visual Form

Monocular Criteria of Depth. The third or depth dimension of visual space is determined in adult experience by both monocular and binocular stimulus factors. However, the latter are ordinarily regarded as primary, and the former as secondary criteria of depth. This may mean either that the monocular factors act by *conditioning* the binocular ones or that, in general, they are less cogent than are the latter. The monocular criteria of depth include all those features of the stimulus pattern, presented to a single eye, which are normally associated with distance or distance relations. Thus, the appearance of bluish haze, due to atmosphere, is ordinarily significant of distance; the covering of one object by another has a similar meaning; as objects retreat from the eye, their bases normally come closer to the horizon; the directions of shadows, when the position of the light source is known, tell something about depth distances. Motion of the observer also provides a powerful cue to depth because of the variation of the apparent angular velocity of an object as a function of distance from the eye.

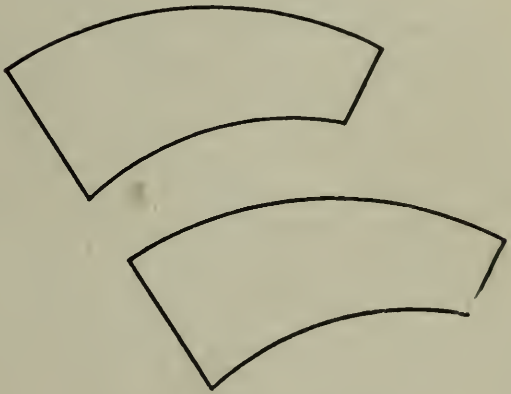
Binocular Criteria of Depth. The binocular criteria of depth rest upon the fact that the projections of identical object fields (having objects at various distances) upon the right and left eyes are never exactly similar, on account of the difference in point of view of the two eyes. However, this dissimilarity, or *disparation,* decreases rapidly with increasing depth,



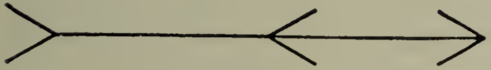
Hering Figure



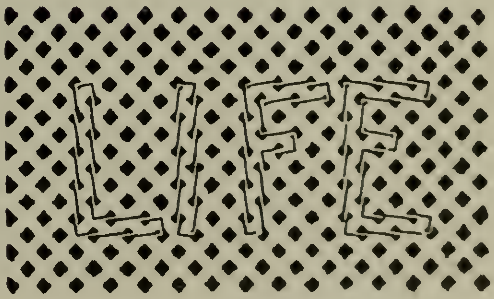
Zöllner Figure



Illusion of Area (Jastrow)



Müller-Lyer Illusion



Twisted Cord Illusion (Fraser)

FIGURE 10

Typical Optical Illusions

(From: Ladd, G. T., & Woodworth, R. S. Elements of physiological psychology. New York: Scribner's, 1911.)

falling below the binocular acuity limit at a distance of about 100 meters. We can, therefore, assume an effective identity of the two retinal patterns for distances beyond this point, with an increasing degree of disparation for diminishing depth distance. Disparation can be measured in terms of the degree of failure of identical object points to be imaged upon *corresponding retinal points.* The latter are defined as comprising the pairs of points which would coincide if the two retinas were brought into superposition, with their physiological centers and vertical axes in registration. When the eyes are in their normal positions, the axes, corresponding to the individual monocular apparent verticals, make an angle of about two and a quarter degrees with each other, converging downwardly to a vertex which is approximately in the floor plane (22, vol. 3, pp. 424-426). The lines of regard, passing through the respective retinal centers, meet at the object point which is fixated.

Horopters. Under such conditions the projections of certain pairs of corresponding points, in addition to the central ones, intersect geometrically in objective space, and the locus of such intersections is known as an *horopter.* It is clear that there must be a different horopter for each separate position of the lines of regard. When the latter meet at infinity, the horopter is a plane surface approximately coinciding with the floor plane, but for nearer positions of convergence, with the fixation point in the median plane, the horopter consists of a circle and a straight line. The circle passes through the point of fixation and the nodal points of the two eyes, while the line lies in the median plane and intersects the circle at the fixation point. The horopter, for any given setting of fixation, furnishes a theoretical reference system with respect to which disparation can be measured. Object points lying on the horopter are imaged on corresponding retinal points and have zero disparation. All other object points, however, will be disparately represented, and to a degree which is greater, the greater their removal from the horopteral locus. The disparation will vary in sign as "crossed" or "uncrossed," according as the distance is greater or less than that of the horopter.

Disparation has two quite different kinds of consequences in consciousness: (*a*) depth localization, and (*b*) double-imaging or *diplopia.* The latter appears when a certain threshold degree of disparation is exceeded and is relatively uncommon or unnoticed except under experimental conditions. We may consider that a tendency towards diplopia always exists whenever identical object points are not imaged upon corresponding retinal points but that this tendency is normally transformed into a depth impression which varies in magnitude and sign with the degree and direction of the latent diplopia.

General Theory of Visual Space. The theory of disparation provides us with a doctrine of localization relative to the horopter, but does not unequivocally relate the subjectively represented position of the latter, for various fixation points, to kinaesthetic or tactual space. Experiments directed towards demonstrating a basis for such localization in sensations of accommodation or of convergence have given negative results (55). It, therefore, appears that the coordination of visual with kinaesthetic depth is probably accomplished through the medium of relations between the visual and kinaesthetic perception of parts of the body other than the eyes, and to an accumulation of memory records of experiences obtained in moving about in space. The total structure of visual space in the depth dimension must be admitted to depend in a composite way upon a large number of very different factors. Thus, although disparation criteria are usually capable of overcoming secondary criteria which conflict with them, this is not always the case. The instrument known as the pseudoscope, which places the image patterns belonging to the right eye upon the retina of the left eye, and vice versa, brings about an inversion of most objects as regards convexity or concavity but is unable to accomplish this result for the human face.

IX. The Visual Perception of Motion

Visual motion differs from that of physical objects in that the former is absolute whereas the latter is always relative. Although visual motion ordinarily involves a visual "thing" which undergoes displacement, motion sensation seems to occur in some instances without there being any definite moving thing or any accumulative displacement. The phenomenon of isolated motion has been designated by Wertheimer (74) by the symbol *phi*. The stimulus to visual motion may consist either in the "real" motion of a physical object or it may lie in a series of discrete images presented discontinuously, although in rapid succession, both in space and in time. The eyes may either pursue the stimulus or may remain stationary. In the latter case, the moving thing in consciousness corresponds to the image which is displaced across the retinal receptors, but in the former case the moving thing is related to a retinally stationary image. It will, therefore, be appreciated that the facts of visual motion perception are so complicated as to preclude a thorough discussion in the present chapter.

Thresholds for Motion Perception. The velocity of the stimulus and the length of path moved through are best expressed in visual angular terms, although visual movement is ordinarily represented in three dimensions, and may be very rapid even with low angular speeds. When the angular velocity is less than one-half to one minute of arc per second near the fixation point, motion impression fails. This threshold increases as the stimulus is displaced towards the periphery of the visual field, becoming 34 minutes of arc per second at 20 degrees (6). At very high speeds of the stimulus, in the neighborhood of 2000 degrees per second, over a path of 20 degrees, there is also failure of motion perception, due to persistence of vision, which yields a stationary streak instead of a moving spot or thing. The threshold for *length of path* is about 17 seconds of arc, under optimal conditions, being lower for rapid than for slow motions. This measure of *motion acuity* decreases with displacement of the stimulus towards the periphery, but less rapidly than does static acuity. Motion acuity is decreased by a factor of about one-quarter when stationary comparison objects are entirely absent. The above thresholds hold under the condition that the eyes do not pursue the moving stimulus. Pursuit by the eyes reduces apparent speeds by a factor of about one-half.

The Stroboscopic Illusion. The possibility of obtaining a motion impression, in the absence of a continuously moving object, by applying a succession of discretely displaced stimuli, is implied by the finite value of motion acuity. However, an impression of smooth motion can be obtained with discontinuous stimuli which are separated by angular distances more than 300 times as great as the minimal path threshold, showing that (cortical) factors other than motion acuity are involved. The most important conditions, determining the "stroboscopic illusion," are the time interval and the spatial (angular) separation between successive exposures. Increasing the time interval from a very small value yields a series of qualitatively different phenomena, as follows: (1) with very small intervals, simultaneous presentation or overlap of successive phases; (2) "part-motion," involving disappearance of the apparently moving thing midway between the initial and terminal positions; (3) smooth or "good" movement; (4) part motion, again, or possibly a pure "phi phenomenon"; and (5) successive appearance of two distinct things without motion.

Using angular separations of about one degree, Wertheimer (74, p. 168ff.) found simultaneity at 30σ, good motion at 60σ, and discrete succession at 200σ or greater, with individual phase exposures lying between 5 and 30σ; Lincke (38) obtained motion impression up to 600σ. When the angular separation is increased, the motion may become jerky and, under some circumstances, tends to divorce itself subjectively from the initial and terminal patterns, constituting the so-called "phi phenomenon." According to Wertheimer, this is a pure phenomenon of motion, whereas Dimmick (13) finds it to consist in a "gray flash." Increased separation decreases the range of time interval, between phases, within which smooth motion can be secured. Marbe (40) was able to obtain a good motion impression with a separation of 1.14 degrees and 4.55 exposures per second. The limiting separation for smooth motion is about 4.5 degrees, and no motion impression at all can be obtained above 12.5 degrees.

Motion Impression and Eye Movement. We have noted that visual motion is absolute in the sense that, at any moment, certain features in the subjective visual field are moving, whereas others may be stationary. Their status in this respect can be reversed only by a radical change in the form of the experience, such as occurs in the familiar "railway station illusion," and its correction. Although motion impressions are uniformly aroused when stimulus patterns are displaced across the mosaic of retinal receptors, this process does not determine whether the pattern or its background shall be represented as in absolute motion. Voluntary pursuit movements of the eyes seem to render the background visually stationary, although its image moves across the retina. (On kinds of eye movements, see 17.) The involuntary ocular pursuit movements of nystagmus, however, throw the entire system of visual things into rotation. Voluntary glancing movements of the eyes have no subjective movement consequences and they seem to be accompanied by momentary visual anaesthesia (25).

The conditions for absolute visual movement cannot even be stated by assuming the bodily kinaesthetic system as a stationary reference axis, since in walking or riding the whole bodily complex is seen to be displaced with reference to a stationary world. When the external world is very imperfectly represented, illusions of motion frequently appear. Thus, in looking at clouds drifting across the moon, the former are often given as stationary, with the moon in motion. An isolated bright spot against a wholly dark field shows random drifting movements—the "autokinetic sensation"—even when it is quite accurately fixated. Another very interesting illusion of motion consists in the "visual movement after-image," a streaming effect which is presented on restricted areas of the visual field after these same areas have been subjected, rather continuously, to kinetic stimulation. This after-image is opposite in direction to the original visual movement and can be neutralized by a moving stimulus having the correct opposed velocity.

X. Visual Experience and the Cerebral Cortex

In the above discussion we have been concerned almost exclusively with the correlations holding between stimulus variables and visual experience. However, it seems quite certain that such correlations are actually of an *indirect* type and that the experience in its totality is determined directly by processes within the cerebral cortex. The cortical mechanism itself is very complex and is undoubtedly divided into successive stages, all of which are of importance for the determination of the final visual experience. The ultimate or direct basis of the experience lies, perhaps, in one of the cortical association areas, such as the frontal. The general principles of psychophysics lead us to suppose that each attribute, aspect, or part of visual experience must have a corresponding feature of which it is privately a function in the ultimate cortical process. Frontal lobe disturbances sometimes seriously upset the entire visual system, producing micropsia and other perceptual distortions. We know very little concerning the details of these ultimate relationships.

Relation of Vision to the Calcarine Cortex. However, we have a good deal of positive information concerning the relation of visual experience to the structures and functions of the primary cortical receiving station for optic nerve impulses, the calcarine region of the occipital lobe. It appears in the first place that the retina, and hence the subjective visual field, has a point-to-point or part-to-part representation on the calcarine surfaces, so that a certain pattern of excitation here is associated with definite retinal and experiential configurations. The left halves of both retinas (corresponding to the right half of the visual field) are represented on the left occipital lobe, while the right retinal halves are connected with the right occipital lobe. The central or foveal region, however, sends connections to both lobes, thus having a double representation for safety (44).

Complete destruction of these occipital areas produces total blindness, while pathological degenerations result in weaknesses and disturbances of vision, varying widely in nature in accordance with the character of the lesion. Local cortical destructions or changes yield local blind-spots or "scotomata" in the visual field. If the whole of one lobe is rendered non-functional, blindness in the opposite half of the visual field, for both eyes—a condition known as hemianopia—results. Such artificial blind areas behave similarly to the natural blind-spot; they are not noticed by the patient or subject until he attempts to perceive detail with the affected visual regions. There is either a complete absence of experience for such regions or they are filled in on the basis of outlying types of excitation. Sometimes hallucinatory presentations are given in scotomatic areas.

At the present time very little that is positive can be said concerning the relations of particular visual attributes or dimensions to the cortical process. The calcarine region evidently contains neural mechanisms for the synthesis of similar impulses received from the two eyes; it is the seat of the primary activities of binocular discrimination. That it contains processes corresponding specifically to brilliance and chroma, we can scarcely doubt. Pathological changes in the occipital cortex can effect chroma independently of brilliance and may yield highly specific chromatic effects, such as red vision or erythropsia. The cortex seems to carry out a series of neurological operations upon the incoming optic nerve-currents, gradually putting them into shape for the coordinated control of efferent cortical innervations and the determination of the visual experience. Such operations include the introduction of memory records, which seem to be carried largely in the so-called visuo-psychic areas, immediately surrounding the primary projection zones.

REFERENCES

1. Allen, F. On the discovery of four transition points in the spectrum and the primary colour sensations. *Phil. Mag.*, 1919, **38**, 55-81.
2. Ames, A., Jr., & Proctor, C. A. Dioptrics of the eye. *J. Opt. Soc. Amer.*, 1921, **5**, 22-84.
3. Andrade, E. N. da C. The structure of the atom. London: Bell, 1924. Pp. 314.
4. Baird, J. W. The color sensitivity of the peripheral retina. Washington: Carnegie Institution, 1905. Pp. 80.
5. Bills, M. A. The lag of visual sensation in its relation to wave lengths and intensity of light. *Psychol. Monog.*, 1920, **28**, No. 5. Pp. 101.
6. Bourdon, B. La perception visuelle de l'espace. Paris: Schleicher, 1920. Pp. 442.
7. Broca, A., & Sulzer, D. La sensation lumineuse en fonction du temps. *C. r. Acad. sci.*, 1902, **134**, 831-834, 1903, **137**, 944-946; 977-979; 1046-1049.
8. Bühler, K. Handbuch der Psychologie. Teil 1. Die Struktur der Wahrnehmungen. Heft 1. Die Erscheinungsweisen der Farben. Jena: Fischer, 1922. Pp. 28-35.
9. Cobb, P. W. Photometric considerations pertaining to visual stimuli. *Psychol. Rev.*, 1916, **23**, 71-88.
10. ————. Vision and the brightness of surroundings. *Trans. Illum. Eng. Soc.*, 1916, **11**, 386.

11. Coblentz, W. W., & Emerson, W. B. The luminous radiation from a black body and the mechanical equivalent of light. *Bull. Bur. Stand.*, 1917, **14**, 255-266.

12. Comstock, D. F., & Troland, L. T. The nature of matter and electricity. an outline of modern views. New York: Van Nostrand, 1917. Pp. 203.

13. Dimmick, F. L. An experimental study of visual movement and the phi phenomenon. *Amer. J. Psychol.*, 1920, **31**, 317-332.

14. ————. Note on the series of blacks, grays and whites. *Psychol. Rev.*, 1925, **32**, 334-336. See also: *Amer. J. Psychol.*, 1920, **31**, 301-302.

15. ————. A reinterpretation of the color pyramid. *Psychol. Rev.*, 1929, **36**, 83-90.

16. Dittmers, F. Ueber die Abhängigkeit der Unterschiedsschwelle für Helligkeiten von der antagonistischen Induktion. *Zsch. f. Sinnesphysiol.*, 1920, **51**, 214-232.

17. Dodge, R. Five types of eye movement. *Amer. J. Physiol.*, 1903, **8**, 307-329.

18. Ferree, C. E., & Rand, G. Chromatic thresholds of sensation from center to periphery of the retina and their bearing on color theory. *Psychol. Rev.*, 1919, **26**, 16-41; 150-163.

19. Gertz, H. Ueber autoptische Wahrnehmung der Sehthätigkeit der Netzhaut. *Skand. Arch. f. Physiol.*, 1907, **19**, 381-408; 1909, **21**, 315-350.

20. Hecht, S. On the binocular fusion of colors and its relation to theories of color vision. *Proc. Nat. Acad. Sci.*, 1928, **14**, 237-241.

21. Hecht, S., & Williams, R. E. The visibility of monochromatic radiation and the absorption spectrum of visual purple. *J. Gen. Physiol.*, 1922, **5**, 1-33.

22. Helmholtz, H. L. F. von. Treatise on physiological optics. Rochester, N. Y.: Opt. Soc. Amer., 1924-1925.

23. Hering, E. Grundzüge der Lehre vom Lichtsinn. Leipzig: Engelmann, 1920. P. 294.

24. Hess, C., & Pretori, H. Messende Untersuchungen über die Gesetzmässigkeit des simultanen Helligkeitskontrastes. *Arch. f. Ophth.*, 1894, **40**, 1-24.

25. Holt, E. B. Eye-movement and central anaesthesia. *Psychol. Rev., Monog. Suppl.*, 1903, **4**, No. 17. Pp. 3-45.

26. Ives, H. E. Studies in the photometry of lights of different colors. II. Spectral luminosity curves by the method of critical frequency. *Phil. Mag.*, 1912, 352-370.

27. ————. Studies in the photometry of lights of different colors. IV. The addition of luminosities of different color. *Phil. Mag.*, 1912, 845-853.

28. ————. The resolution of mixed colors by differential visual diffusivity. *Phil. Mag.*, 1918, **35**, 413-421.

29. Ives, H. E., & Kingsbury, E. F. The theory of the flicker photometer. *Phil. Mag.*, 1914, **28**, 708-728; 1917, **34**, 99-112.

30. Jaensch, E. R. Ueber den Farbenkontrast und die sogennante Berücksichtigung der farbigen Beleuchtung. *Zsch. f. Sinnesphysiol.*, 1921, **52**, 165-180.

31. Jones, L. A. The fundamental scale of pure hue and retinal sensibility to hue differences. *J. Opt. Soc. Amer.*, 1917, **1**, 63-77.

32. Jones, L. A., & Lowry, E. M. Retinal sensibility to saturation differences. *J. Opt. Soc. Amer.*, 1926, **13**, 25-34.

33. Katz, D. Die Erscheinungsweisen der Farben, u. s. w. *Zsch. f. Psychol. Erg.*, 1911, **7**. Pp. xviii+425.

34. Kirschmann, A. Ueber die quantitativen Verhältnisse des simultanen Helligkeits und Farben-Contrastes. *Phil. Stud.*, 1891, **6**, 417-491.

35. König, A., & Dieterici, C. Die Grundempfindungen in normalen und anormalen Farbensysteme und ihre Intensitätsverteilung im Spektrum. *Zsch. f. Psychol.*, 1892, **4**, 241-347.

36. Ladd-Franklin, C. The nature of the colour sensations. In Helmholtz's Treatise on Physiological Optics. Rochester, N. Y.: Opt. Soc. Amer., 1924, 1925.

37. ————. Radiation visible provenant de fibres nerveuses excitées. *C. r. Acad. sci.*, 1927, **185**, 584-587.

38. Lincke, P. Neue stroboskopische Versuche. *Ber. d. II. Kong. f. exper. Psychol.*, 1907, 214-219.

39. Mach, E. Contributions to the analysis of the sensations. (Trans. by C. M. Williams.) Chicago: London, 1897. Pp. 208.

40. Marbe, K. Theorie der kinematographischen Projecktionenen. Leipzig: Barth, 1910. Pp. 80.

41. Martin, M. F. Film, surface and bulky colors and their intermediates. *Amer. J. Psychol.*, 1922, **33**, 451-480.

42. McDougall, W. The sensations excited by a single momentary stimulation of the eye. *Brit. J. Psychol.*, 1904, **1**, 78-113.

43. ————. The variation of the intensity of visual sensation with the duration of the stimulus. *Brit. J. Psychol.*, 1904, **1**, 151-189.

44. Morax, V. Discussion des hypothèses faites sur les connexions corticales des faiseaux maculaires. *Ann. d'oculist*, 1919, **156**, 25-35.

45. Nutting, P. G. 1919 report of standards committee on visual sensitometry. *J. Opt. Soc. Amer.*, 1920, **4**, 55-79.

46. Parsons, J. H. An introduction to the study of color vision. (2nd ed.) Cambridge: Cambridge Univ. Press, 1924.

47. Piéron, H. De la variation de l'énergie liminaire en fonction de la surface rétinienne excitée pour la vision periphérique (Cônes et bâtonnets). *C. r. Soc. biol.*, 1920, **83**, 753-755.

48. ————. De la variation de l'énergie liminaire en fonction de la durée d'excitation pour la vision fovéale. *C. r. Acad. sci.*, 1920, **170**, 525-528.

49. Preston, T. The theory of light. (4th ed.) London: Joly, 1901. Pp. 586.

50. Priest, I. G. Preliminary note on the relation between the quality of color and the spectral distribution of light in the stimulus. *J. Opt. Soc. Amer.*, 1920, **4**, 388-401.

51. ————. Note on the relation between the frequencies of complementary hues. *J. Opt. Soc. Amer.*, 1920, **4**, 402-404; **5**, 513.

52. Purkinje, J. E. Beobachtungen und Versuche zur Physiologie der Sinnesorgane. Beiträge zur Kenntniss des Sehens in subjektiver Hinsicht. Prag, 1819. Neure Beiträge u. s. w., Berlin, 1825.

53. Quain, J. Anatomy. (Ed. by E. A. Schäfer *et al.*) New York, 1909, Vol. 3, Part 2. Pp. 185-220.

54. Reeves, P. The rate of pupillary dilation and contraction. *Psychol. Rev.*, 1918, **25**, 330-340.

55. Rivers, W. H. R. Vision. Schäfer's Text-Book of Physiology, II, 1132-1135. Edinburgh, London: Pentland, 1898-1900.

56. Roelofs, C. O. Ueber Wettstreit und Schwankungen im Sehfelde. *Arch. f. Ophth.*, 1921, **104**, 133-148.

57. Russell, H. N. The minimum radiation visually perceptible. *Astrophys. J.*, 1917, **45**, 60-64.

58. Schjelderup, H. K. Ueber eine vom Simultankontrast verschiedene Wechselwirkung der Sehfeldstellen. *Zsch. f. Sinnesphysiol.*, 1920, **51**, 176-213.

59. ————. Zur Theorie der Farbenempfindungen. *Zsch. f. Sinnesphysiol.*, 1920, **51**, 19-45.

60. Sherrington, C. S. The integrative action of the nervous system. New York: Scribner's, 1906. Pp. 411.

61. Sinden, R. H. Studies based on the spectral complementaries. *J. Opt. Soc. Amer.,* 1923, **7**, 1123-1153.

62. Smith, F. O. An experimental study of retinal sensitivity and discrimination for purple under different degrees of intensity of stimulation. *J. Exper. Psychol.,* 1925, **8**, 381-397.

63. Spearman, C. The new psychology of shape. *Brit. J. Psychol.,* 1924, **15**, 211-225.

64. Titchener, E. B. A text-book of psychology. New York: Macmillan, 1909. Pp. 565.

65. Troland, L. T. Report of the Colorimetry Committee of the Optical Society of America. *J. Opt. Soc. Amer.,* 1922, **6**, 531-535.

66. ————. The measurement of visual stimulation intensities. *J. Exper. Psychol.,* 1917, **2**, 1-33.

67. ————. The retinal visibility function. *Abstr., Bull. Phys. Lab. Nat. Elec. Lamp Asso.,* 1, No. 3, 378-382.

68. ————. The colors produced by equilibrium photopic adaptation. *J. Exper. Psychol.,* 1921, **4**, 344-390.

69. ————. The influence of minuthesis due to one spectral stimulus upon the luminosity of another. *Abstr., Bull. Phys. Lab. Nat. Elec. Lamp. Asso.,* 1, No. 3, 393-399.

70. ————. Apparent brightness: its conditions and properties. *Trans. Illum. Eng. Soc.,* 1916, 964-965.

71. ————. Preliminary note: The influence of changes of illumination upon after-images. *Amer. J. Psychol.,* 1917, **28**, 497-503.

72. Troland, L. T., & Langford, C. H. The stability of hue under chromatic minuthesis. *J. Opt. Soc. Amer.,* 1925, **10**, 294.

73. Voeste, H. Messende Versuche über die Qualitätsänderungen der spectralfarben in Folge von Ermüdung der Netzhaut. *Zsch. f. Psychol.,* 1898, **18**, 257-267.

74. Wertheimer, M. Experimentelle Studien über das Sehen von Bewegungen. *Zsch. f. Psychol.,* 1912, **61**, 161-265.

75. Westphal, H. Unmittelbare Bestimmungen der Urfarben. *Zsch. f. Sinnesphysiol.,* 1909, **44**, 182-230.

76. Zoth, O. Augenbewegungen und Gesichtswahrnehmungen. In Nagel's Handbuch der Physiologie des Menschen. Braunschweig: Vieweg, 1905.

CHAPTER 5

VISION: II. THE NATURE OF THE PHOTORECEPTOR PROCESS

SELIG HECHT
Columbia University

I. SIMILARITIES OF SENSE-ORGANS

The environment in which animals normally live may be considered as constant in its broadest aspects. In its details, however, the environment is always changing. It is with this diversity and its evaluation that the sensory economy of animals is concerned. Animals do not accomplish this evaluation exhaustively. Of the numerous components of our surroundings with which we have become familiar through modern physics and chemistry, only a portion can be received directly through the established sensory channels. However, the variety is large enough to cover what appear to be the more common changes in the environment.

The different sense-organs which have become established are not necessarily concerned with qualitatively different aspects of environment, even though the sensations which result in man may be qualitatively different. Thus the light sense and the temperature sense are both concerned with radiations which differ only in wave-length. Similarly, smell and taste are both concerned with physicochemical changes which differ apparently only in magnitude of concentration. Nevertheless, different sense-organs do record portions or phases of the environmental flux, which differ significantly even if not specifically.

It is apparent that, in order to achieve this diversification in the evaluation of the outside world, sense-organs must differ one from another. There has to be a specific affinity between a particular type of physical or chemical disturbance and the sense-organ concerned with recording this disturbance. Such a specialized correlation cannot be absolute, because a sense-organ is a living tissue, and as such it is subject to the limitations of the properties of living material. For example, so highly individualized a sense-organ as the eye is brought into play in its specific way by a number of heterologous stimuli as varied as a mechanical blow and an electric shock, largely because all living material is affected by these agents. Nevertheless, the receiving end of a sense-organ must be of a configuration corresponding to the particular type of environmental change which it registers.

It is hardly necessary to labor this point because of the very obvious structural differences which sense-organs show. What needs to be emphasized is the idea that sense-organs have more in common than merely the philosophical distinction of giving us information about the environment. Sense-organs are a group of structures which possess certain fundamentally similar properties. For example, all sense-organs possess a threshold. This is the minimum intensity of the outside environmental change which on prolonged exposure will just produce an effective activity of the sense-organ. The presence of a threshold is not exclusively a sense-organ property but is possessed as well by nerve, by muscle, and by many other living structures. It is, however, an important characteristic of sensory phenomena. Another property of sense-organs is the capacity for adaptation. A third and perhaps the most useful sensory property is the ability to discriminate intensities. There are many other characteristics common to sense-organs which one might enumerate, but enough has been said to indicate that different sensory systems are functionally similar. It would be surprising if they were very different in physicochemical organization.

A sense-organ possesses an essentially dual function. On the one hand it receives an environmental disturbance, and on the other it transmits this information to an attached nerve-fiber. The reception of the outside changes requires a specialized arrangement in the sensory process suited to the nature of the environmental alteration, because this is specific. The kinetics and dynamics of this relationship, as represented by the threshold, adaptation, and the like, are, however, quite similar in different sense-organs. And the final activity of the sensory process, the initiation of a nerve-impulse, is also similar. The last twenty years of research into the physiology of nerve has shown that the nerve-impulse is essentially the same from nerve to nerve, even from motor to sensory nerve. It is, therefore, likely that the method of eliciting this impulse is also the same from sense-organ to sense-organ. The initiation of a nerve-impulse is probably accomplished normally by the change in potential at a given region of the nerve. This may then be the final thing which all sense-organs do at their transmitting end.

We now have the first generalization which may be made about the receptor process. Speaking functionally, its outside end is a special configuration which can be upset by a particular part of the environment; its middle part is not dissimilar from one sense-organ to another, and concerns the quantitative interrelations of environment and sensory process; its inside end produces the same result in all sense-organs, namely, a nervous impulse.

In the following pages I propose to describe what appear to be the fundamental properties of the receptor process, to study the data pertaining to them, and then to see whether it is possible to suggest a type of physicochemical organization in terms of which these properties find a rational expression. In the detailed treatment I shall use almost exclusively the process of photoreception. However, from what has already been said it is apparent that while the minute analysis of one sensory response does not presuppose an identical mechanism in another, the basic ideas derived are of general interest.

II. Threshold

An environmental disturbance which starts a receptor process can vary in three ways: in intensity, in time, and in composition. Let us leave composition to one side for a moment, since I shall devote a separate section to it later, and consider first the relationships which exist between time and intensity.

In general, in order for an outside agent such as light to produce a given sensory effect, the intensity required is an inverse function of the duration of its application. The exact relation between the two is still unsettled in its final form for the human eye and it varies to some extent from organism to organism; but in all cases it is true that the lower the intensity, the longer is the time necessary for it to act, and, conversely, the higher the intensity, the shorter is the time required to elicit an effect.

Such a relationship is a commonplace of photochemistry. In its simplest form it is the Bunsen-Roscoe law (15) which holds for many photochemical reactions; it states that to produce a given photolytic effect a constant amount of energy is necessary, regardless of its distribution in time. If I is the intensity and t the time of its action then

$$It = C \qquad [1]$$

where C is a constant. As here given, or in a modified form the inverse relationship applies to a great variety of photolytic processes.

The relation between time and intensity in the photoreceptor process brings to light a significant property of vision, namely, the threshold. For a given animal and under constant conditions there can be found a low, but measurable intensity which produces an effect after a reasonably long exposure, but below which even an extremely prolonged exposure fails to elicit a sensory effect. This intensity is the threshold.

What does the presence of an intensity threshold tell us about the photosensory process? The production of a given sensory effect undoubtedly means the formation of a certain physical or chemical change in the contents of the sense-cell. In the photosensory process the first effect of the light is obviously the conversion of a light-sensitive but inactive substance into an active substance which starts the train of events culminating in a nervous discharge. The first thing then that we learn from the existence of a threshold intensity is that this primary photochemical reaction is accompanied by another reaction, not photochemical in nature, which tends to remove the products of the photochemical reaction. If the first part of the photoreceptor process were merely a simple, direct, photochemical reaction, we should never have a threshold. The reciprocity law of Bunsen and Roscoe would probably hold rigorously, as it does in the irreversible bleaching of visual purple (32) among other things, and exposures would be cumulative as they are in the photographic plate. Like the astronomer who can expose a photographic plate to light of the same dim star several evenings in succession and get a developable image, we should be able to see a light no matter how low its intensity, provided the exposure were prolonged sufficiently. The effects of light, however, do not accumulate this way in the sense-cells. Therefore, there must be a process which tends to remove or render inactive the photochemical products as they are formed by the light.

One may put the whole matter in the following way. Let us describe the photosensitive material in the sense-cell as S and suppose that light changes S into its photochemical product, which may be one substance P, or more substances P, A, B, etc. We may write $S \rightarrow P$ as this reaction. It is P which starts the series of events ending in the outgoing impulse to the nerve. The associated process demanded by the existence of a threshold may then be of the following kind. P may diffuse away from the scene of action. P may combine with some constituent of the cell and be rendered inactive. Or, finally, P, by itself or by combining with some material furnished by the cell or by the surrounding fluids, may become reconverted into the original sensitive material S and be used over again.

On the basis of the information so far presented, one cannot decide which of these conditions is the most probable. But for purposes of understanding the data which describe the properties of the threshold they may be treated mathematically in a practically identical manner. Assume an intensity I shining on such a system. The light will be absorbed by the substance S, which will be changed from its initial concentration a to a concentration $a—x$ where x is the concentration of S converted into P. Since this is a threshold change, it will be small, and we may write its average velocity as

$$v_1 = k_1 I \ (a - x) \qquad [2]$$

where k_1 is a velocity constant which includes the absorption coefficient. Now assume that P diffuses away, or combines with another substance, or is changed back to S. The velocity of the process will be proportional to the concentration of the reactants, $P, A, B \ldots$, which in case of chemical reactions would be x, or x^2, or x^n depending on whether there are 1, 2, or n number of different photolytic products. This velocity

$$v_2 = k_2 \ x^n \qquad [3]$$

must be subtracted from the photochemical velocity in order to describe the actual velocity

$$\frac{x}{t} = k_1 I \ (a - x) \ - k_2 x^n \qquad [4]$$

with which the products P accumulate in the sense-cell. We are not interested particularly in this velocity, but in the relations between the intensity I and the time t which are required to result in a given constant amount x of photochemical products accumulating in the sense-organ. Equation [4] may be arranged as follows:

$$It = \frac{x}{k_1 \ (a - x)} + \frac{k_2 x^n}{k_1 \ (a - x)} \, t \qquad [5]$$

to give us the desired information. Since x is taken as a constant amount of material, $(a - x)$ is also constant, and equation [5] may be written

$$It = C + Dt \tag*{[6]}$$

where C and D are constants.

Equation [6] has been derived somewhat differently by Lasareff (56) and by Pütter (71), but, aside from its derivation, it and its predecessor equation [5] have many interesting properties. The velocity constant k_2 of the negative reaction is most likely smaller than the velocity constant k_1 of the photochemical reaction; otherwise it would require enormous intensities to effect any measurable accumulation of products P in the sense-cell. Thus the constant D is probably smaller than C. Moreover, for threshold values x is a small fraction of a; any power of x must therefore be smaller than x, which makes the value of D still smaller. Finally, with small values of the exposure time t, the term Dt in [6] becomes negligible in relation to C, and the equation becomes

$$It = C \tag*{[1]}$$

which is the Bunsen-Roscoe law all over again. What this means in words is that, in spite of the presence of the extra reaction which removes P, we might expect the relation between time and intensity in the photosensory system to follow the very simple equation [1] for short exposures, and the slightly more involved equation [6] for longer exposures. This is essentially what has been found by a variety of investigators.

Bloch (10) first found that for *short flashes* of light the relation between I and t for the human eye conformed to equation [1]. This has been confirmed by Rutenberg (76) and Braunstein (13). Table 1 gives Braunstein's data from which the results may be judged. A similar application of the Bunsen-Roscoe law has been found (23) for short flashes of light in the ascidian *Ciona intestinalis*, the clam *Mya arenaria*, and by Loeb (61) for a small range of intensities with the polyp *Eudendrium*. In Table 1 there is included the data for *Mya*. In some organisms the relation $It = C$ holds over such an enormous range of time and intensity that it is fairly certain there is no accessory reaction to remove the photolytic products as they are formed. Blaauw (8) showed for wheat seedlings that a given phototropic response is caused by the same amount of energy, about 20 meter-candle-seconds, when the exposure time varies between

TABLE 1

RELATION BETWEEN INTENSITY I AND TIME t FOR SHORT EXPOSURES TO PRODUCE A MINIMAL SENSORY RESPONSE IN *Mya* AND IN THE HUMAN EYE

Data for the eye are from Braunstein (13); for *Mya* from Hecht (23).

	Eye			*Mya*	
I	*t*	*I.t*	*I*	*t*	*I.t*
1	.0210	.021	54	0.104	5.62
2	.0098	.020	76	0.073	5.55
4	.0047	.019	112	0.053	5.94
8	.0024	.019	194	0.030	5.82
16	.0012	.019	238	0.023	5.47
			334	0.016	5.34

0.001 second and 43 hours, and the intensity between 26,520 and 0.00017 meter-candles. Over such a range this seems never to have been found with an animal.

When the exposures become longer than the short flashes already considered, it has been shown by a number of observers that equation [6] applies. Blondel and Rey (11) found that for exposures between 0.001 and 3.0 seconds their data for the human eye may be expressed by the equation $It=0.065+0.31t$, using arbitrary values of the intensity. This relation has been confirmed by various experimenters. For extremely short exposures it seems not to hold, as both Piéron (69) and Reeves (74) have shown. Apparently other factors enter. In the case of *Mya,* Piéron (68) believed that for a larger range of exposures than those given in Table 1 the simple It relation does not hold. Unfortunately Piéron's experiments are vitiated by his use of an episcotister for varying the intensity; since this device produces a series of short flashes, it obviously cannot be used for short exposures; and, furthermore, there is implicit in it the assumption that I and t are interchangeable, which is the question at issue. Curiously enough, for the longer exposures, where one might expect an episcotister to be useful, Piéron's data show that for *Mya* equation [6] applies and that $It=3.5+1.2t$.

There is an interesting consequence of the difference in the relation between time and intensity for short and for long exposures. The physicochemical reason that It is constant for short exposures may be that, since the straight photochemical reaction is so rapid, the comparatively slow removal process hardly gets under way before the sensory response is over. Under such conditions the initial process may be treated as a purely photochemical reaction. It has become almost axiomatic to say, since Goldberg (22) first showed it to be true, that photochemical reactions possess low temperature coefficients very near 1 for 10°C., in contrast to ordinary chemical reactions whose temperature coefficients for 10°C. are between 2 and 3. For short flashes of light it should therefore be true that the value of It obtained for the photosensory process at different temperatures should yield a coefficient near unity. Using a constant exposure of 0.016 second, I determined the minimum intensity necessary to elicit a response from *Mya* at four temperatures between 13° and 31°C. The results (23) with four separate experiments gave values for the temperature coefficient for 10°C. as follows: 1.04, 1.06, 1.07, and 1.06, which bear out the idea that for short flashes the first reaction is primarily photochemical.

These examples show that the existence of a threshold, if interpreted even in the most general terms, is enough to furnish information about the structure of the receptor process. The assumption of a removal or recovery process accompanying the principal, photochemical reaction has enabled us to derive equations which describe as a first approximation the known data of the relation between time and intensity in the visual process.

The characteristic of the threshold which is to be emphasized is that intensity is the limiting factor, and not time. Below a certain intensity, no matter how long the exposure may be, no sensory effect is produced.

Consider this in terms of equation [4] describing the velocity of the reaction as a whole. If the reaction proceeds for infinite time, $t = \infty$ and x/t becomes 0. Equation [4] then becomes

$$\frac{k_1}{k_2} I = \frac{x^n}{a - x} \qquad [7]$$

from which time is completely eliminated. A given concentration x of P can be formed, according to this equation, only as the result of a certain intensity I, even at infinite exposure.

The conception of the photoreceptor process which we have so far suggested has been deliberately general. It has been specific in the one respect only, that it has involved the existence of a process which accompanies the main photochemical process and removes its products. It is possible to be even more specific about this reaction, but in order to do so we must consider other characteristics of the photoreceptor process and the data that belong to them.

III. Duality of Sensory Process

It is a curious but significant fact that we see a light only after it has been on for some time, and continue to see it after it has been turned off. In the human eye, the period during which the light must be on and the period during which it may be off are very short and require special devices for their demonstration. Nevertheless, the cinematograph bears witness to their existence, and the literature of after-images to their fascination. This property of photoreception is well distributed among other animals, and in some it is so apparent that it may be demonstrated and studied with comparative ease. It may be well to describe the responses of such an animal in order to make the rest of the treatment in this chapter more intelligible.

If an animal like *Mya,* whose picture is shown in Figure 1, is exposed to light, it responds by a vigorous retraction of its extended siphon. The time from the beginning of the exposure to the beginning of the retraction is called the reaction-time, and is a constant quantity under constant conditions. This reaction-time is obviously not a simple period, since it includes the time of action of the light, the other reactions in the sensory cells, the conduction of the impulses, and the latent period of the muscle response in the siphon. In the case of *Mya* and of such similarly reacting animals as *Pholas* and *Ciona* the reaction-time is of the order of magnitude of three or four seconds. Plainly in a period of this duration it is safe to neglect the very short periods of time occupied by such rapid processes as nerve conduction and muscular latent period. In fact it can be shown by comparison with the reaction-time to mechanical stimuli, which is very short, that the reaction-time to light is taken up almost entirely with the processes in the sensory cells.

The exposures to light which *Mya* requires in order to respond are only a small fraction of the reaction-time, as may be seen from Figure 2. This exposure or sensitization period represents the time for the accumula-

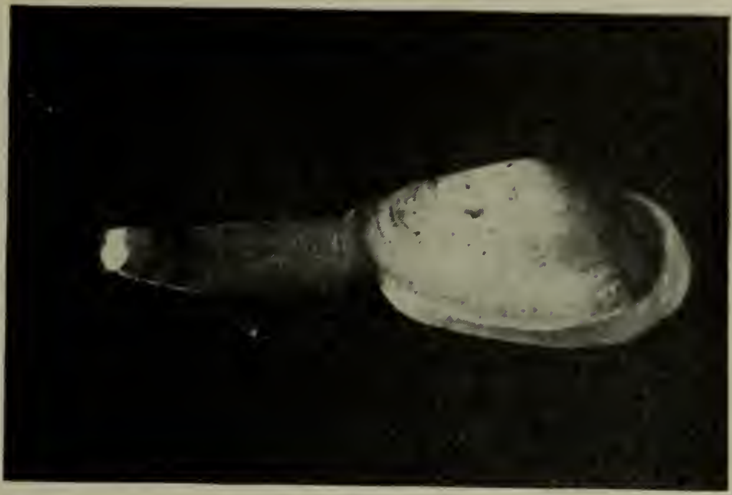


FIGURE 1

Photograph of a Medium-sized *Mya arenaria*, with its Double Siphon Moderately Expanded

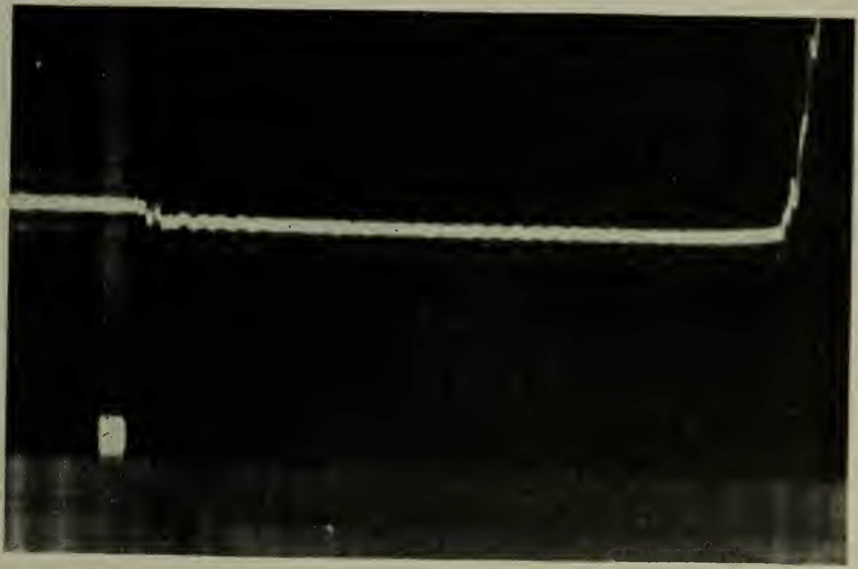


FIGURE 2

Optical Record of Stimulation and Response of *Mya*

The white streak running the length of the figure represents the movement of a lever attached by means of a thread to the siphon tip of *Mya*. The white spot and vertical streak at the left record the duration of the exposure (0.07 second). The siphon begins to retract about 2 seconds after the exposure. The time in between is the latent period. The tuning fork gives 50 vibrations per second. The slight irregularity in the siphon record immediately after the exposure is due to an accidental vibration of the delicately balanced lever caused by the movement of the shutter.

tion in the cell of the necessary concentration of photolytic products formed by the primary photochemical reaction and its associated removal reaction. The greater part of the reaction-time is the latent period, during which the animal need not be in the light in order to respond to it. Most of the duration of the photoreceptor process is thus taken up not by the photochemical reaction, but by some other process which follows it in time.

This secondary, latent period process is an integral part of the receptor mechanism because there is a definite connection between the two parts of the reaction-time. If the exposure of *Mya* or *Ciona* to light is maintained until the animal responds by a siphon retraction, the reaction-time of the animal is of minimum duration. Exposure for shorter intervals, down to a certain point, results in no change in the reaction-time. Exposure below this critical value (sensitization period) causes the resulting reaction-time to increase. This change in total reaction-time is obviously

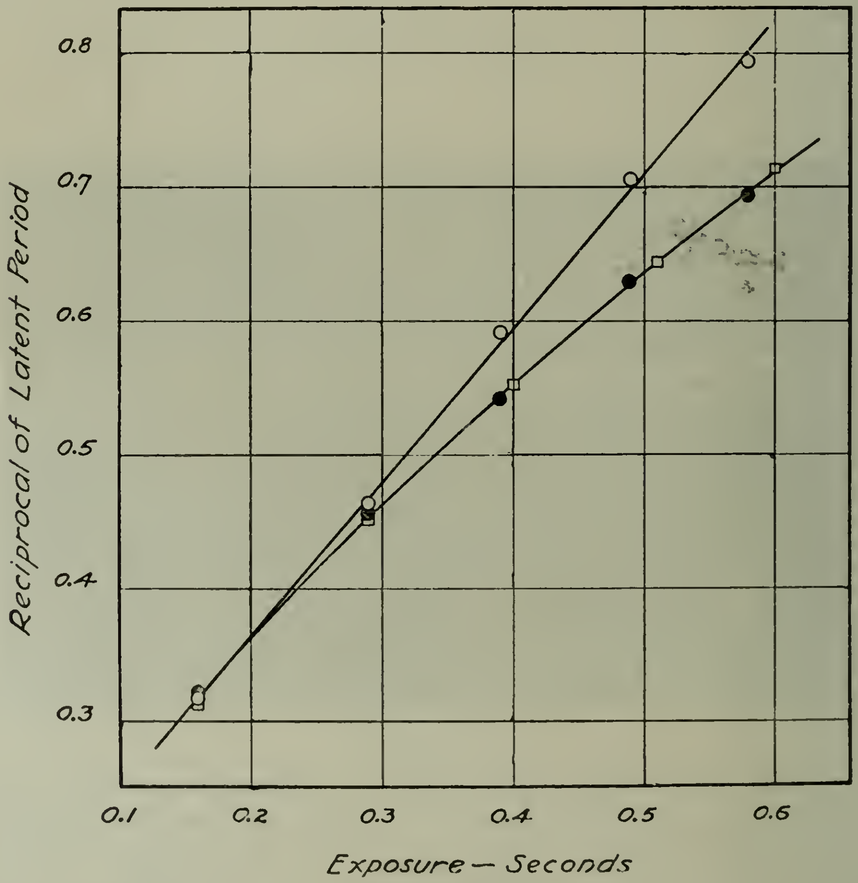


FIGURE 3

Relation between the Duration of the Exposure to a Constant Intensity of Light and the Reciprocal of the Resulting Latent Period for *Ciona*

The connection between the two is practically linear, being exactly so for one series of experiments, and very nearly so for the two other series given in the figure.

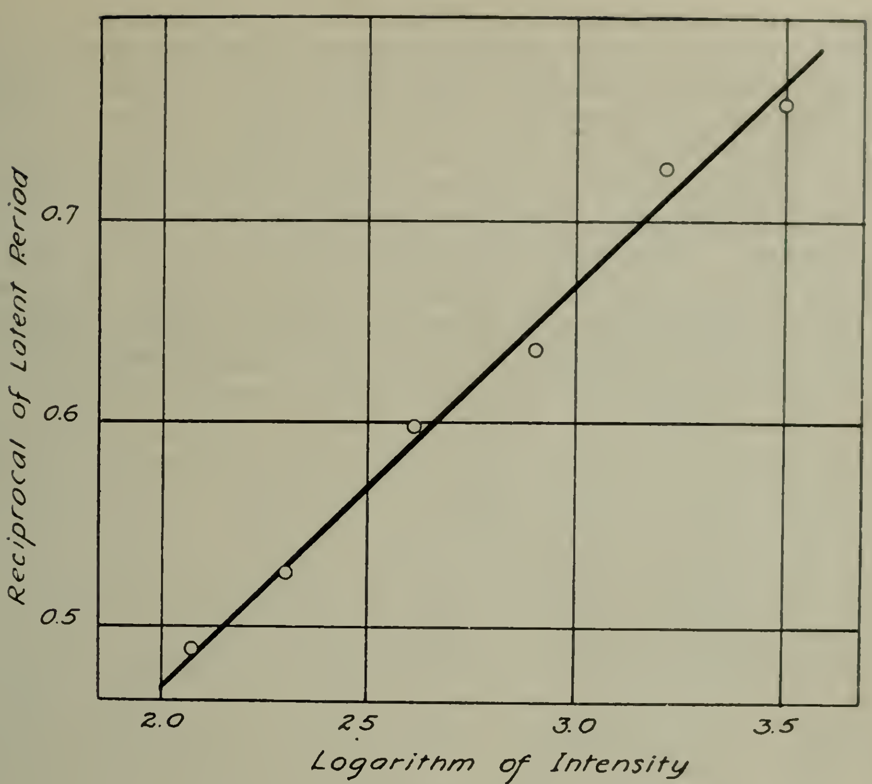


FIGURE 4

RELATION BETWEEN THE INTENSITY OF THE LIGHT AT CONSTANT EXPOSURE AND THE RECIPROCAL OF THE RESULTING LATENT PERIOD FOR *Mya*

The connection between log *I* and the reciprocal of the latent period is obviously linear

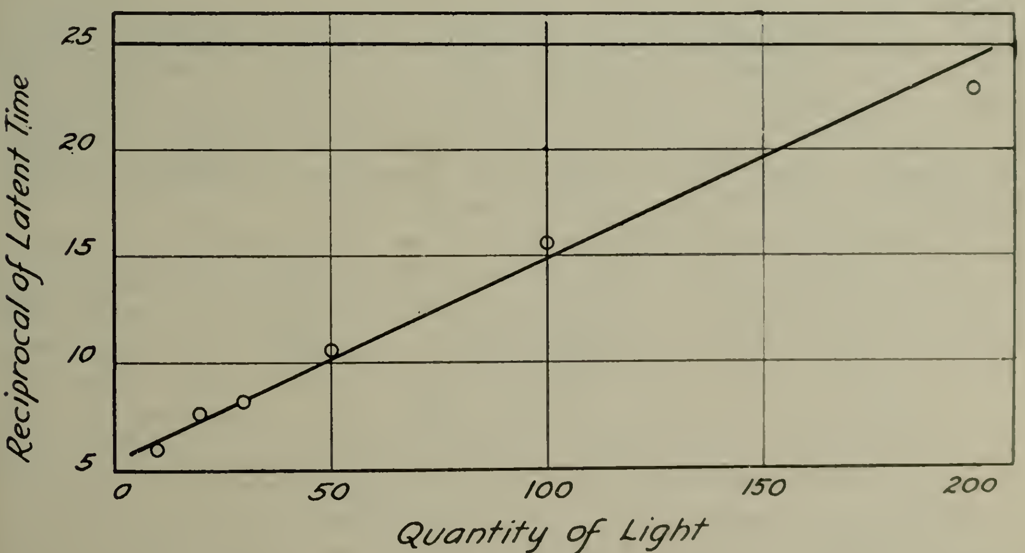


FIGURE 5

RELATION BETWEEN QUANTITY OF LIGHT AND THE RECIPROCAL OF THE RETINAL REACTION-TIME FOR THE EEL'S EYE

Data from Adrian and Matthews (4).

an increase in the latent period. Experiments (28) made to determine the relation between the length of exposure to a constant intensity of light and the duration of the resulting latent period show that the reciprocal of the latent period is very nearly a linear function of the length of the exposure period. The results of some experiments of this nature with *Ciona* are given in Figure 3. For these short exposures to light one may assume that the photochemical effect of the light is directly proportional to the exposure. These data then tell us that the velocity of the secondary, latent period process (which equals the reciprocal of the latent period) varies directly with the concentration of photolytic products formed during the exposure period. From this we learn that during the latent period something happens whose velocity depends on the chemical effect produced by the stimulating light.

If this is really so, then some other method of varying the photochemical effect should produce similar results. In the above experiments the intensity is kept constant and the exposure varied. If now the exposure period is kept small and constant, and the intensity of the light is varied, it is found, as Figure 4 shows, that the velocity of the latent period process varies directly with the logarithm of the intensity. It is known from much photochemical work, particularly with the photographic plate, that for constant exposures the effect of light varies as the logarithm of its intensity. This confirms our idea that the latent period represents a process whose speed varies with the concentration of photolytic products formed in the primary reaction of the receptor process. It may be added that, if the experiments are so arranged that both time and intensity are varied, the expected results appear, namely, that the reciprocal of the latent period varies directly with the product of the time and the logarithm of the intensity (29).

These experiments can be performed with *Mya* and *Ciona* with a stop-watch and a photographic shutter. However, the relationships between the two parts of the receptor process hold even when the periods of time require more elaborate methods of control and measurement. Figure 5 is a record of one of the many experiments by means of which Adrian (4) has recently shown that essentially the same type of behavior may be demonstrated in the retina of the vertebrate eye.

The significance of these relationships for an understanding of the second part of the receptor process lies in the fact that they show that the velocity of the process which takes place during this period varies directly with the concentration of photochemical products produced during the primary exposure period. The possible interpretations of these relationships will help us understand the sensory process. The primary photochemical effect may be thought of as occurring at one end of a sense-cell, and the setting off of the nerve at the other end. The photolytic substance P at one end would then have to diffuse over, and the latent period would represent the duration of this diffusion. The fairly exact linear relation between concentration and latent period velocity renders this diffusion idea somewhat improbable because the rate of diffusion is a linear function of the concentration of a diffusing substance only when it diffuses into a large volume where its concentration may be regarded as zero. This is obviously not the situation here since it is essential for the substances to accumulate in sufficient quantity to initiate a response in the nerve. A more likely interpretation is that the latent period represents the duration of a chemical reaction which is essential for the process. This reaction is all set and ready to go, but must have as a component some substance which is produced by the photochemical reaction. This may be something which catalyzes the second reaction, or it may actually combine with the second reaction and become an integral part of it. In either case, the accumulation of a definite amount of the active products of the latent period reaction discharges chemically or electrically the impulse which goes to the attached nerve and starts the nervous end of the process.

In the case of some animals it has been possible to make a choice experimentally between the notion that the latent period represents a diffusion process and the notion that it is a chemical reaction. This is done by studying the influence of temperature on the duration of the latent period, because of the strikingly different effects of temperature in the two cases.

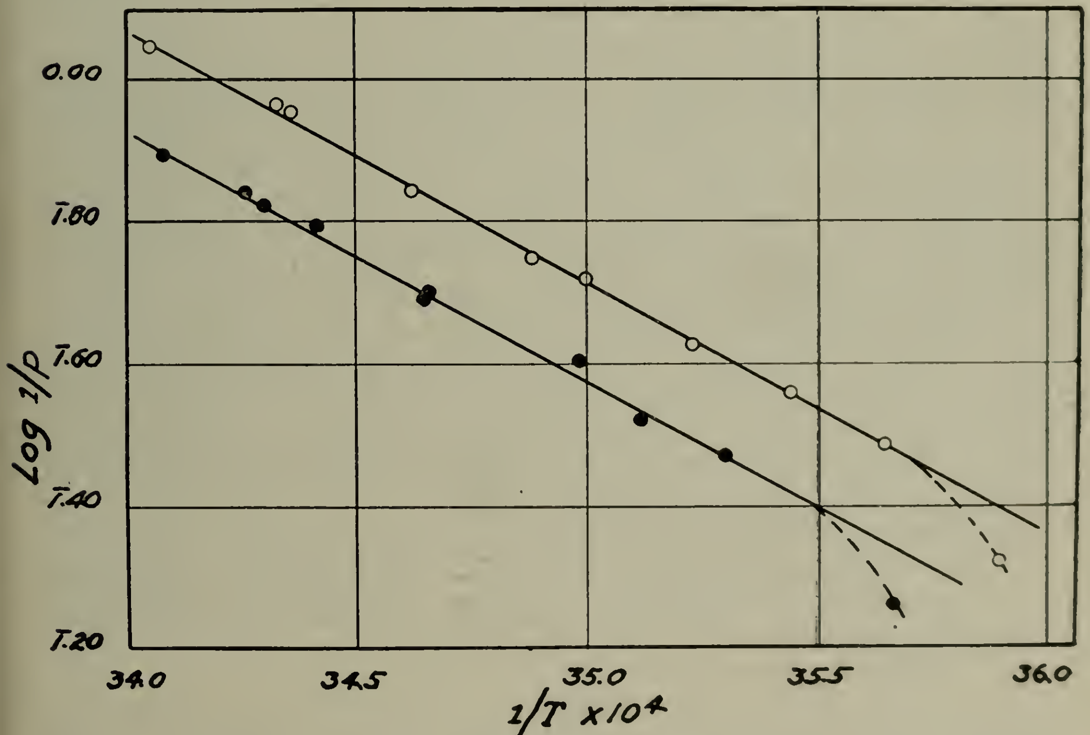


FIGURE 6

Relation between the Absolute Temperature (T) and the Latent Period (P) in *Ciona*

The points are from two series of experiments, whereas the two straight lines are computed from the Arrhenius equation $\log 1/p = -\mu/RT + C$ where $\mu = 16{,}200$. In the equation R is the gas constant, and C is an arbitrary integration constant of no significance here.

The effect of temperature on the velocity of diffusion processes is negligible, whereas its effect on chemical reactions is considerable. Experiments made with *Mya, Ciona,* and *Pholas* (28) show that the latent period velocity increases with temperature. Figure 6 gives the results of such an experiment. The velocity of the latent period turns out to be an exponential function of the absolute temperature and follows the classical Arrhenius equation which has been used in chemistry to describe the influence of temperature on chemical reactions. This is true in all the three species in which this relation has been investigated.

The second part of the receptor process is therefore undoubtedly a chemical reaction. Exactly what type of reaction, it is hard to say. The values of μ in the Arrhenius equation describing the reaction (cf. Table 2) would indicate that it is not the same in all animals. Whether it is a reaction demanding a catalyst like the photoproducts P or whether it actually involves combination with one of these products cannot be decided now. Either relation fits all that we know so far. In fact, Adrian's work on the vertebrate retina (4) indicates that the relation between the primary photochemical reaction and the secondary, "dark," latent period reaction may be neither of these, but of still another kind. Further work is clearly necessary.

Before concluding this section it would be well to put together in a formal manner what we have so far learned about the photoreceptor process. If we call L an indifferent substance present in the sense-cell, we may suppose that under definite conditions it is converted into a chemically or electrically active product T, a given concentration of which sets off the attached nerve. The duration of the reaction $L \rightarrow T$ constitutes the latent period. However, the reaction $L \rightarrow T$ though all set and ready to go cannot do so unless there is present in the cell the photolytic products formed from the sensitive material S, which either catalyze the process or combine with L to form T. If we represent this relation between the photochemical products and the latent period reaction in a conventional manner as $||P||$, then the receptor process may be diagrammatically represented as $S \rightarrow P;\ L\ ||P|| \rightarrow T$, the first reaction of which is controlled by light and the second uninfluenced by light.

There is omitted for the moment the fact that associated with the reaction $S \rightarrow P$ is one which removes P or renders it inactive. I shall make up for this omission by treating this reaction in considerable detail in the following sections.

TABLE 2

VALUES OF μ FOR THE EFFECT OF TEMPERATURE ON THE PHOTOSENSORY LATENT PERIOD OF DIFFERENT ANIMALS

Species	μ
Ciona intestinalis	16,200
Mya arenaria	19,700
Pholas dactylus	18,300

IV. Light-Adaptation

Nearly all animals can become adapted to light by means of their photoreceptors. For example, after we have been in the dark for several hours, even a comparatively weak illumination represents a strong stimulus, as the common experience of turning on a light in the middle of the night bears witness. In a short time, however, our eyes become indifferent to this illumination, so much so that we move on the street during the day surrounded by illuminations many thousand times more intense than the weak light which annoys us when it is suddenly turned on in the dark.

A similar thing is true of other animals, for example, the clam *Mya*. If after being in the dark for some time it is suddenly illuminated, it responds to this illumination by a vigorous retraction of its siphons. If the illumination is permitted to continue, the animal no longer responds to it. Again to elicit the characteristic response of the organism, it is necessary to increase the intensity of the light. The necessary extent of this increase varies with the intensity of light (26) to which the animal is already adapted, and may be considered a measure of the animal's sensitivity. Light-adaptation thus means a loss in sensitivity, since more and more light becomes necessary to elicit the minimal response.

From this description of light-adaptation two things are apparent; first, that an animal can come into sensory equilibrium with its environment and maintain this state indefinitely; and second, that the condition is reversible. Both these characteristics of adaptation are of importance in understanding the nature of the photoreceptor process. Consider what the continued maintenance of a given condition of sensory equilibrium means in terms of what we learned in the previous sections. If the sensitive material S is exposed continuously to light of a given intensity, it will be broken down and used in initiating impulses. The sensory effect of this decomposition is limited, as we saw, by the speed with which the accompanying reaction inactivates or removes the photochemical products P from the cell. It is apparent, however, that if this is to be maintained indefinitely there must be a continuous supply of S furnished to the sense-cell. Otherwise all the available S would be used up by the light in a short time. This would not be adaptation by the maintenance of an equilibrium, but the simple exhaustion of a supply. Therefore, as it is being used up by the light, S must be replenished in some way.

The reversibility of adaptation points to the same conclusion. Normally we become light-adapted daily through a whole series of illuminations and then dark-adapted at night—a process which goes on all our lives. Experimentally this process may be repeated many more times, and it is found that the sensitivity of the eye decreases with light-adaptation and recovers with dark-adaptation over and over again. In other words, a continuous supply of S to the sense-cell is essential and must be considered as a part of the photosensory process.

As in the case of the reaction which removes the photochemical products P, there are several possibilities for this restorative process. There may be plenty of S or its main constituent in the blood stream, which then supplies it to the sense-cell by diffusion. In this way the concentration of *S* would be practically constant. Or it may be that the main photochemical products recombine among themselves or with some other substance furnished by the cell in excess, like water or oxygen, and form anew the sensitive material from which they were formed. One cannot make final choice between these possibilities, but there exists evidence which renders the latter idea more acceptable. This evidence depends on the information which is furnished by the study of dark-adaptation.

V. DARK-ADAPTATION

The increase in sensibility of the eye in the dark has been familiar since it was first described by Aubert (6). After a half hour's stay in the dark the sensitivity of the human eye may increase to as much as a hundred thousand times its original condition in daylight. The course of this change is surprisingly regular, as the many measurements by Piper (70), Nagel (66), Kohlrausch (51), and myself (25) have shown. That the same things are true of other animals is of comparatively recent knowledge (33). If the course of dark-adaptation is followed in such animals as *Mya, Ciona, Pholas,* or the frog tadpole (67), the data show a uniform type of behavior, which I shall present first because of its comparative simplicity. An animal is exposed to strong light for some time, after which it is placed in the dark at a given moment. At different times in the dark its reaction-time to a light of constant intensity is measured. It is found that this reaction-time decreases steadily until after several hours it reaches a constant minimum characteristic of its response to that intensity. The reaction-time when plotted against time in the dark falls on a smooth curve. Figure 7 shows the course of dark-adaptation in *Ciona;* the data for *Mya* and the other species resemble these with *Ciona.*

This much is fact. The interpretation to be derived from it sheds much light on the nature of the underlying process of photoreception. The reaction-time of *Ciona* to light is occupied by several processes which combine to produce a constant effect, namely, a siphon retraction. The algebraic sum of the velocities of these separate processes is represented by the reciprocal of the reaction-time. As we saw before, only two processes, the photochemical and the latent period, need be considered. The data then show that the sum of these velocities increases during dark-adaptation. In order to interpret the quantitative way in which dark-adaptation proceeds, we need to make but one assumption. This is that since the end result of each measurement is always the same, namely, a siphon contraction, the photochemical effect necessary to produce it is also always the same. We have already learned that the velocity of the latent period is directly proportional to the photochemical effect. Since the photochemical effect is assumed constant, it follows that the velocity of the latent period is also constant. The total velocity of the reaction-time processes is thus composed of two velocities, one of which is constant. Any increase in total velocity must therefore represent an increase in the velocity of the primary photochemical process only; and hence the curve in Figure 7

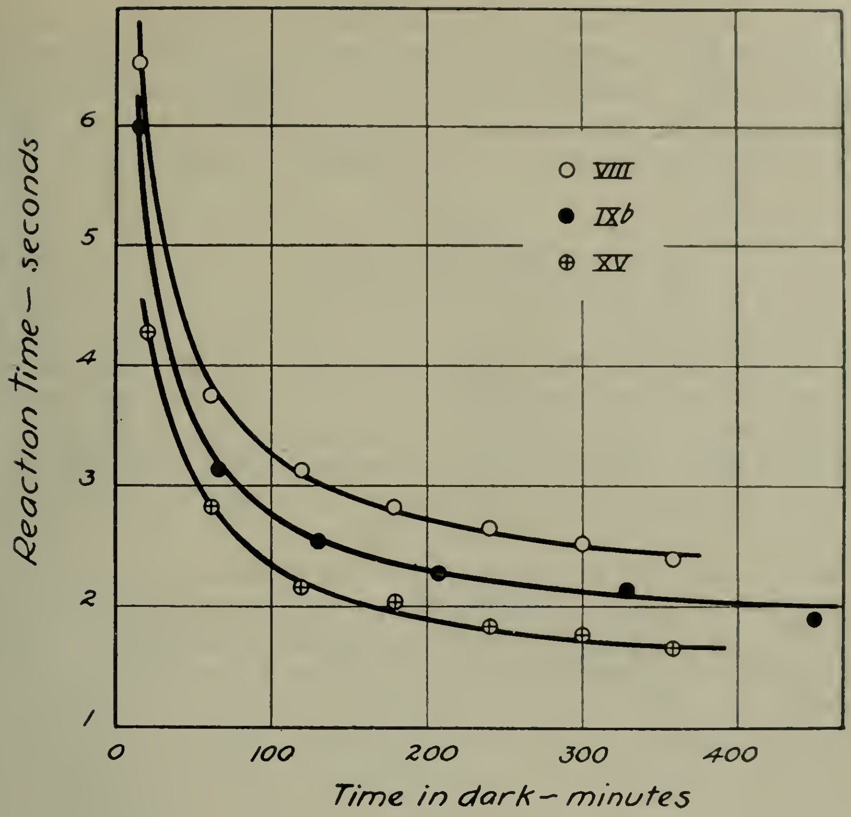


FIGURE 7

Dark-Adaptation of *Ciona*

The data represent three experiments; the points for XV have been lowered 0.5 second in the drawing in order to keep them distinct from the others. The curves drawn through the points are all three computed in terms of a bimolecular reaction as explained in the text.

represents the changes which take place in the velocity of the photochemical reaction required to produce a sensory response during dark-adaptation.

Since the intensity of the measuring light is constant, the differences in velocity of the primary photochemical reaction are very likely due to differences in the concentration of sensitive substance. On obvious photochemical and mass law grounds, it may be supposed that the velocity of the reaction is proportional to the concentration of photosensitive substance in the sense-cells. It then follows that the concentration of sensitive material in the sense-cell increases during dark-adaptation in conformity with the shape of the experimental curve of dark-adaptation. When the form of the curve in Figure 7 is investigated, it is found to be that of a bimolecular reaction isotherm

$$k_2 = \frac{a - x}{t\, a\, x} \qquad [8]$$

where k_2 is the velocity constant; t the time of dark-adaptation; $a - x$ the concentration of photosensitive material already formed; and x the concentration still to be formed.

Actually the numerical values are secured in this way. If we call r_0 the reaction-time at the beginning of dark-adaptation, r_t the reaction-time at any moment t during dark-adaptation, and r_f the final reaction-time after complete dark-adaptation, then $a = 1/r_f - 1/r_0$; $a - x = 1/r_t - 1/r_0$; and $x = 1/r_f - 1r_t$. These values may be substituted in equation [8] which becomes

$$k_2 = \frac{r_f^2 \; (r_0 - r_t)}{t \; (r_0 - r_f) \; (r_t - r_f)} \qquad [9]$$

and into which the experimental values of the reaction-time may be substituted for purposes of computation. It is to be pointed out that the curves drawn in Figure 7 are actually the theoretical curves drawn from equations [8] and [9]. Table 3 gives an example of the numerical agreement of the experimental data with calculation according to the theoretical interpretation.

The conclusion which we may draw from this consideration of the data is that the sensitive material S accumulates in the sense-cells during dark-adaptation as if it were being formed from two other substances. Diffusion usually follows a first-order, not a second-order process. Since at least two substances seem to be concerned in the formation of the sensitive material, they are probably combining in a chemical manner. The chemical nature of the process of dark-adaptation can be tested by studying the effect of temperature on the process of dark-adaptation. We have already seen that the velocity of diffusion processes is usually only slightly affected by a rise in temperature, while that of chemical processes is significantly dependent on the temperature. Experiments were therefore made with *Mya* in which dark-adaptation was measured at three temperatures, about 5°C. apart. The measurements (33) show that dark-adaptation is about three times as rapid at the highest temperature as at the lowest. The

TABLE 3

THEORETICAL ANALYSIS OF THE DARK-ADAPTATION OF *Ciona*. SERIES XV

The reaction-time at complete adaptation is $r_f = 1.82$ seconds.

$k_2 = 0.0268$; $a = 0.426$.

Time in dark	Reaction-time r_t	
	Observed	Computed, $r_t = \dfrac{159.0 + 1.82t}{19.7 + t}$
min.	sec.	sec.
22	4.77	4.77
60	3.35	3.36
120	2.67	2.70
180	2.53	2.44
240	2.31	2.29
300	2.24	2.21
360	2.13	2.14

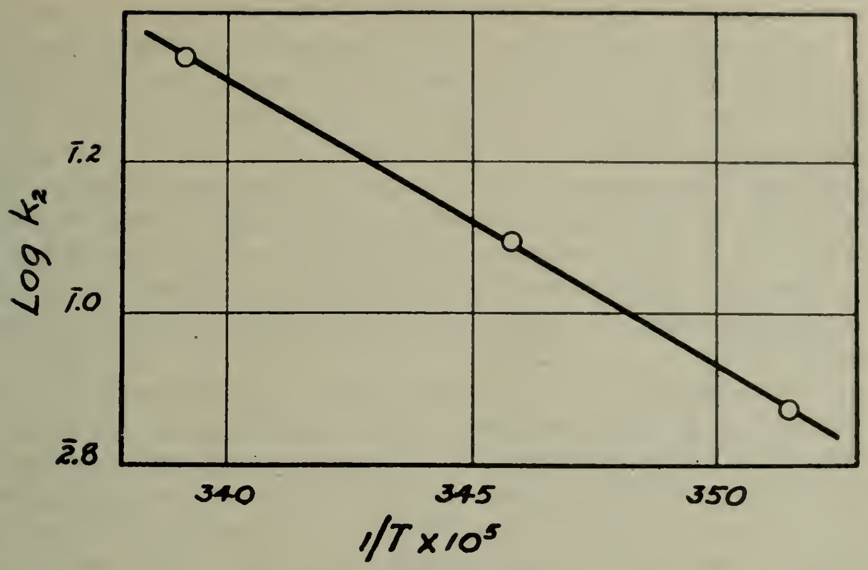


FIGURE 8

Relation between the Absolute Temperature T and the Velocity Constant k_2 of the Dark-Adaptation Reaction in *Mya*

The straight line is the Arrhenius equation $\log k_2 = -\mu\ RT + C$ with $\mu = 17{,}400$.

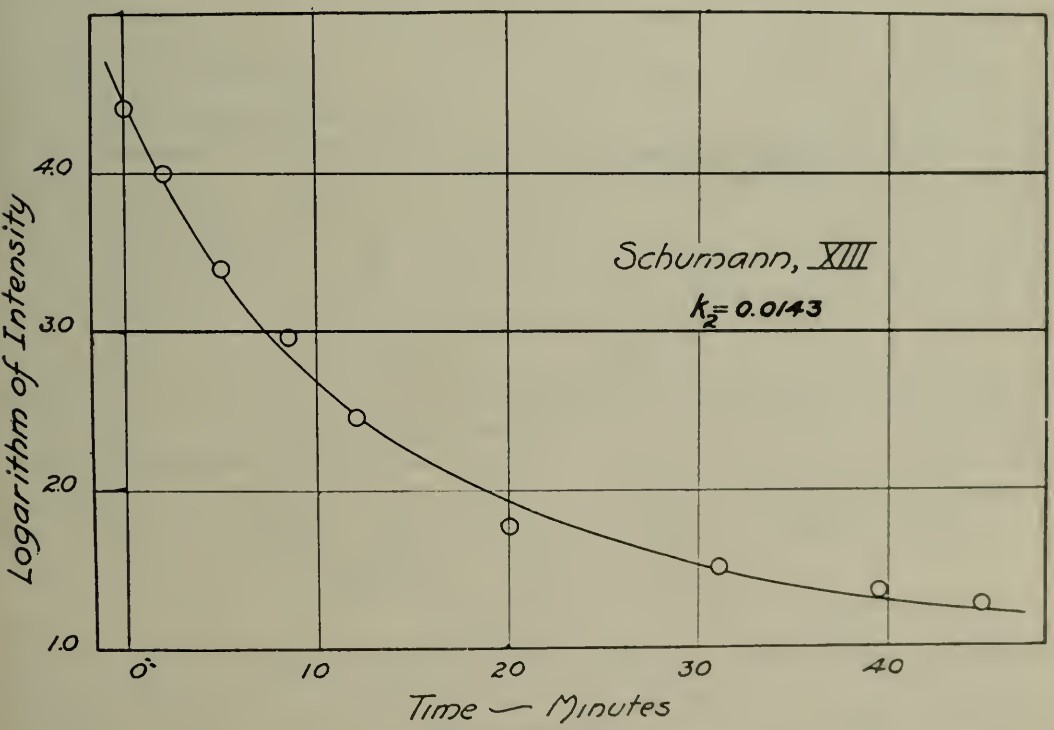


FIGURE 9

Dark-Adaptation of the Human Eye Representing Only the Function of the Rods

Data from Piper (70). The curve is computed from a bimolecular reaction as explained in the text.

data are sufficiently adequate quantitatively so that they may be analyzed in detail. Figure 8 shows that the velocity constant k_2 of the bimolecular equation is related to the absolute temperature according to the Arrhenius equation, which describes such relations in pure chemistry.

It is surprising how adequately the kinetics of a bimolecular reaction describes the dark-adaptation of quite a variety of animals from the clam to the human eye. The human eye is particularly interesting because it is a double sensory system composed of two sets of receptors, rods for dim vision and cones for bright vision involving color. These two systems work quite independently in dark-adaptation and their behavior may be studied separately.

The earlier investigations were concerned almost entirely with the rods. Figure 9 gives an example of the result that was obtained; it is from the work of Piper (70), who made the first reasonably adequate measurements with the rods. The data are plotted as log I following an old suggestion by Best (7). Fifteen years later I (25) made the first complete measurements with the cones. It should have then been apparent that the dark-adaptation of the eye as a whole includes both sets of phenomena; and indeed some of the older measurements show this. However, it was

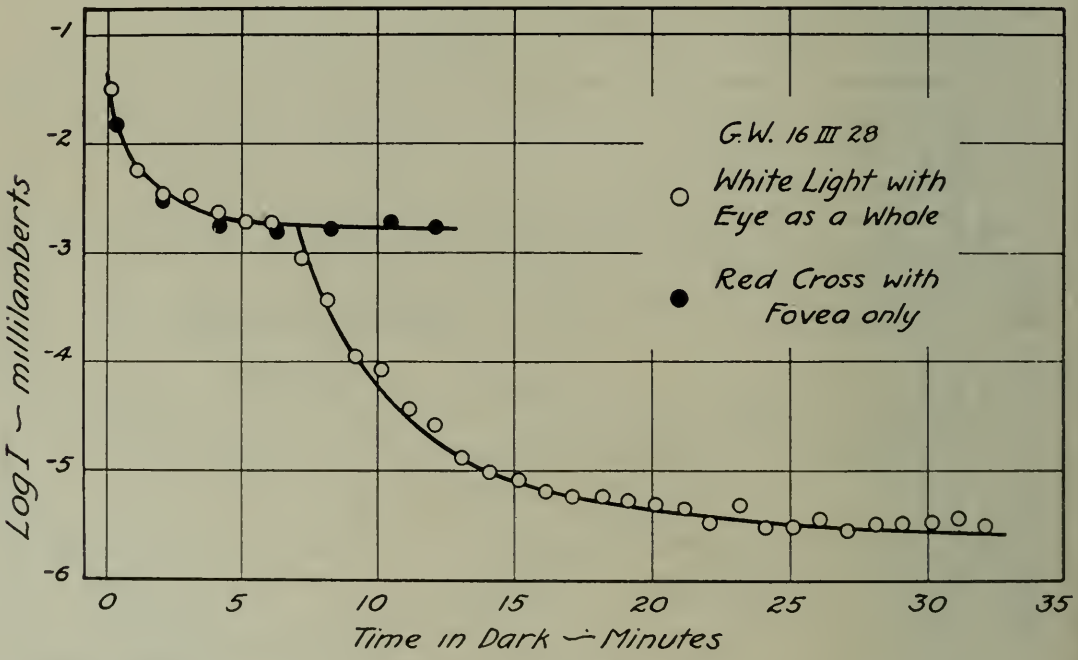


FIGURE 10

Dark-Adaptation of the Eye, as a Whole, Showing the Independent Action of the Cones and Rods

The open circles are one experiment in which the whole eye was measured, whereas the solid circles are another experiment in which the cones in the fovea alone were measured. The upper curve for the eye, as a whole, obviously concerns only the cones; the rest concerns the rods. The curves are theoretical, and represent the kinetics of a bimolecular reaction in each case. The data are measurements made by Messrs. Charles Haig and George Wald in our laboratory.

Kohlrausch (51) who first pointed out that, as Figure 10 demonstrates, the adaptation of the eye as a whole shows a sharp break where the adaptation of the cones practically ceases, and the function is taken over and carried on by the rods alone.

Lasareff (58), Pütter (72), and I (24) have each independently made mathematical analyses of the dark-adaptation process in the eye. While differing in detail and in the accuracy with which they reproduce the data, all these analyses assume that dark-adaptation is an expression of the kinetics of the reaction which restores the photosensitive substance to the sensitive elements. The interpretation which I have just given of the data of animals like *Mya* and *Ciona* shows how reasonable such an idea is. Lasareff's first analyses (58, 56) were made in almost complete disregard of the actual data. His equation describes the known numerical measurements only after 30 minutes of dark-adaptation when, as Figure 9 shows, the process is very nearly complete. As a result of criticism he has recently (59) included the really significant, early part of the course of dark-adaptation. This has led him into complicated considerations of after-images; and the resulting derivations describing the complete process have lost whatever directness and simplicity the earlier equations possessed. Pütter's (72) treatment of dark-adaptation is one phase of an interesting effort (71) to derive all sensory phenomena from certain a priori considerations in themselves not unreasonable. The result is an equation which is almost identical with Lasareff's first equation, and is subject to the same criticism.

The theoretical analysis of the dark-adaptation of the eye which I proposed rests on the experimental fact that in the photosensory process in *Mya* the photochemical effect of light is proportional to the logarithm of the intensity (cf. Figure 4). This relationship is found in many photochemical reactions; and there are so many properties of the eye which are logarithmic functions of the illumination that one may assume this photochemical relation to hold for the eye as well. This means that starting with a constant, initial concentration of sensitive material S, the amount of it which is changed by the light is proportional to the logarithm of the intensity. It therefore follows for the same photochemical system that, if the initial concentration is variable and unknown, it will be proportional to the logarithm of the light intensity necessary to produce a small but constant photochemical effect. The proportionality is inverse, because the lower the concentration, the higher is the intensity which is necessary to produce a constant small effect.

The measurements of dark-adaptation consist of determinations of the intensity of light required to produce a minimal and therefore constant sensory effect. Figures 9 and 10 show that the values of the intensity decrease steadily. On the other hand, the concentration of photosensitive material undoubtedly increases during the process of adaptation. It is possible to relate these two things in order to derive a quantitative expression for the kinetics of the process, by supposing that the concentration of sensitive material at any moment is an inverse logarithmic function of the threshold intensity at that moment. In other words the concentration of S is proportional to log $1/I$, or to its equivalent —log I.

Using the same notation as before, we may call I_0 the intensity threshold at the first moment of dark-adaptation, I_t the intensity at any moment, t, and I_f the final intensity at complete dark-adaptation. The concentrations of the sensitive material at these moments are then proportional to —log I_0, —log I_t, and —log I_f, respectively. The difference between —log I_f and —log I_0 gives the total change in concentration of sensitive material during dark-adaptation. This we may call a, which then becomes $a = \log I_0/I_f$. Similarly the difference between —log I_t and —log I_0 gives the change in concentration of sensitive substance already accomplished at the moment t. This is $a—x$ and becomes $a—x = \log I_0/I_t$. Finally the difference between —log I_f and —log I_t gives the concentration x of sensitive material still to be formed before dark-adaptation is complete. Thus $x = \log I_t/I_f$. Substituting these values in equation [8] we get

$$k_2 = \frac{\log I_0/I_t}{t\ (\log I_0/I_f)\ \ (\log I_t/I_f)} \qquad [10]$$

which may then be used for computing any set of experimental data.

Actually the curve drawn in Figure 9 is the theoretical equation [10] with $k_2 = 0.0143$; and the curves drawn in Figure 10 are those of equation [10] with $k_2 = 0.525$ for the cones, and $k_2 = 0.089$ for the rods. It is apparent that the experimental points are represented with a satisfactory degree of accuracy by the equations. The reason that k_2 for the rods in Figure 9 is different from that in Figure 10 is that the unit of intensity is different in the two cases. One may therefore conclude that during dark-adaptation the sensitive material S is being replenished in the rods and in the cones independently as if in each case it were being formed by the chemical interaction of at least two other substances.

VI. The Photoreceptor Process

We have now arrived at a critical point in the effort to make a physicochemical picture of the photoreceptor process. Certain things seem to be fairly well established. The process as a whole is composed of two parts: a primary photochemical system and a secondary "dark" system. These are intimately related to each other in that the dark process, though all ready to go, cannot proceed to form the necessary material which initiates the nerve-impulse until the primary process has produced some products as the result of the chemical action of light on a photosensitive material. Moreover, the velocity of the secondary dark process is proportional to the concentration of the photo-products formed in the primary reaction.

So far as the primary reaction is concerned, we know three significant things. First, a sensitive material S is decomposed by the incident light, as the first step in the sensory process. Second, a part of the resulting products, which we have called P, are by some means taken away as they are formed so that the full photochemical effect is never transmitted. And third, the photosensitive material S is replenished in the dark and presumably also in the light in a way which indicates its chemical formation from two other substances which disappear in the process. All this information can no longer be left in a vague condition, but must be combined into a concrete picture for experimental and mathematical testing.

Let us begin with the primary part of the process. Dark-adaptation is the reaction which restores the supply of sensitive material which is constantly being depleted by the photochemical reaction. If one supposes that the sensitive material is built up mainly from the products of its photolysis, then its kinetics as a bimolecular reaction is almost the simplest conceivable. This amounts to the assumption that the photosensitive material S is decomposed by light into at least two substances, P and A, which at the same time tend to recombine and form the original material S. We then have for the first part of the receptor process a reversible photochemical system, which may be written as a completely reversible reaction

$$S \xrightleftharpoons[\text{“dark”}]{\text{light}} P+A \qquad [11]$$

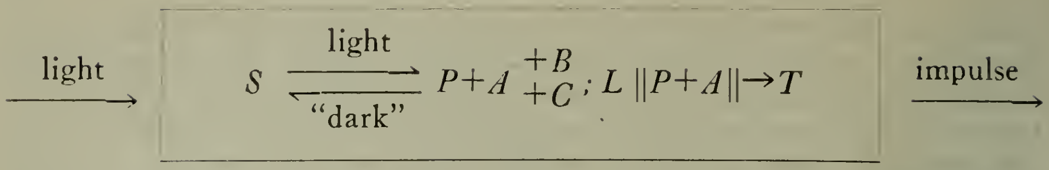
for convenience, but which more likely is only a pseudo-reversible reaction of the type

$$\begin{aligned} S &\xrightarrow{\text{light}} P+A+B \\ S &\xleftarrow[\text{“dark”}]{} P+A+C \end{aligned} \qquad [12]$$

where B is the substance which carries on the effects of the primary light reaction, and C is a substance present in excess which supplies the material and energy for the recombination. The two forms [11] and [12] may be treated mathematically in an essentially identical manner; so for convenience we shall usually refer to the simpler form in equation [11]. But it is well to remember that equation [12] is the more probable form when one tries actually to picture the energetics and material balance which are concerned in photoreception.

For the second part of the receptor process we have seen that $L \rightarrow T$ represents the known facts. The unknown relation between the primary and secondary reaction may be represented by $\|P+A\|$. This may mean catalysis by one or another of the photolytic products of the primary reaction; or it may mean actual combination of L with B formed in the primary process; or it may mean some other interrelation which we cannot think of now.

Keeping all this information in mind, we can then express the physicochemical system which underlies the photoreceptor process in the following diagrammatic form.



$$\xrightarrow{\text{light}} \quad \boxed{S \underset{\text{“dark”}}{\overset{\text{light}}{\rightleftharpoons}} P+A \begin{matrix}+B\\+C\end{matrix} ; L\,\|P+A\|\rightarrow T} \quad \xrightarrow{\text{impulse}}$$

Taken as a whole, this system is a coupled photochemical reaction composed of a primary photochemical reaction and a secondary "dark" reaction. In addition, the primary reaction is reversible. Such coupled photochemical reactions often possess the significant property of involving much more chemical change, both in number of molecules and in quantity of energy, than would be expected from the relatively small amount of light energy used by the primary photochemical reaction. In this respect, it conforms to what one has always supposed the sensory process to be, namely, a nearly self-contained system, high in energy content, which can be released by a small addition of energy from the outside. It is this which is usually meant when the sensory process is spoken of as a trigger mechanism. The energy and reactions are all there; they merely require to be released. In the case of the photosensory process this is done by the formation of decomposition products by the light.

This hypothetical model of the photosensory process must not be taken literally. It is merely a diagrammatic and shorthand summary of the essential characteristics of the sensitivity of animals to light. I have purposely kept it as simple as possible. In the application of mathematics to biology, it seems to me that simplicity is absolutely essential, because even at its best no mathematical treatment can ever be as complex as the organism really is. The simpler a model and the more concrete it is, the more one can think about it in relation to the organism, and the more direct are the experiments that can be made in terms of it. Its deviations from observed fact are then obvious and may be used as the starting-point for new conceptions and fresh experiments. On the other hand, computations based on a complicated model obscure their deficiencies by the closeness with which they may be made to fit the data by the use of many arbitrary constants.

VII. The Chemical Substances

It would be well if one could define more concretely the individual substances and processes which have been schematically represented as composing the photoreceptor system. Only a beginning has, however, been made in this direction. As we saw in Table 2, the values of μ for the relation between temperature and the latent period indicate that the reaction $L \rightarrow T$ is not the same in the three species in which this has been investigated. The organization of the substance is the same, but the substances themselves are undoubtedly different.

This may be shown even better in the identification of the photosensitive substance S. It is well known that the absorption spectrum of a photosensitive substance defines in general its sensitivity toward different portions of the spectrum. Conversely, a study of the sensibility of an animal toward different parts of the spectrum furnishes a means of approximate classification of the spectral characteristics of its photosensitive substance. In recent years as the result of work by Loeb (62), Laurens and Hooker (60), Mast (64), Crozier (16), and myself (29), there has appeared a number of measurements of the most effective portion of the spectrum for the stimulation of different organisms. Although for many animals this point lies in the yellow-green and for many plants in the blue-green, there is an increasing number of organisms which do not conform to this classification. The particular significance of this classification is due to Hess (42). The results of some rather superficially conducted experiments with a number of animals led Hess to the conclusion which he vigorously proclaimed that "the curves of the relative stimulating values of different homogenous (spectral) lights coincide, approximately or exactly, with the luminosity curve of the totally color blind human eye." This he took to mean that these animals are actually color-blind, a conclusion which hardly follows. It might, however, be taken to mean that the sensitive materials of all these different animals are identical. This, however, is also incorrect.

As an example one may take two animals, *Mya arenaria* and *Pholas*

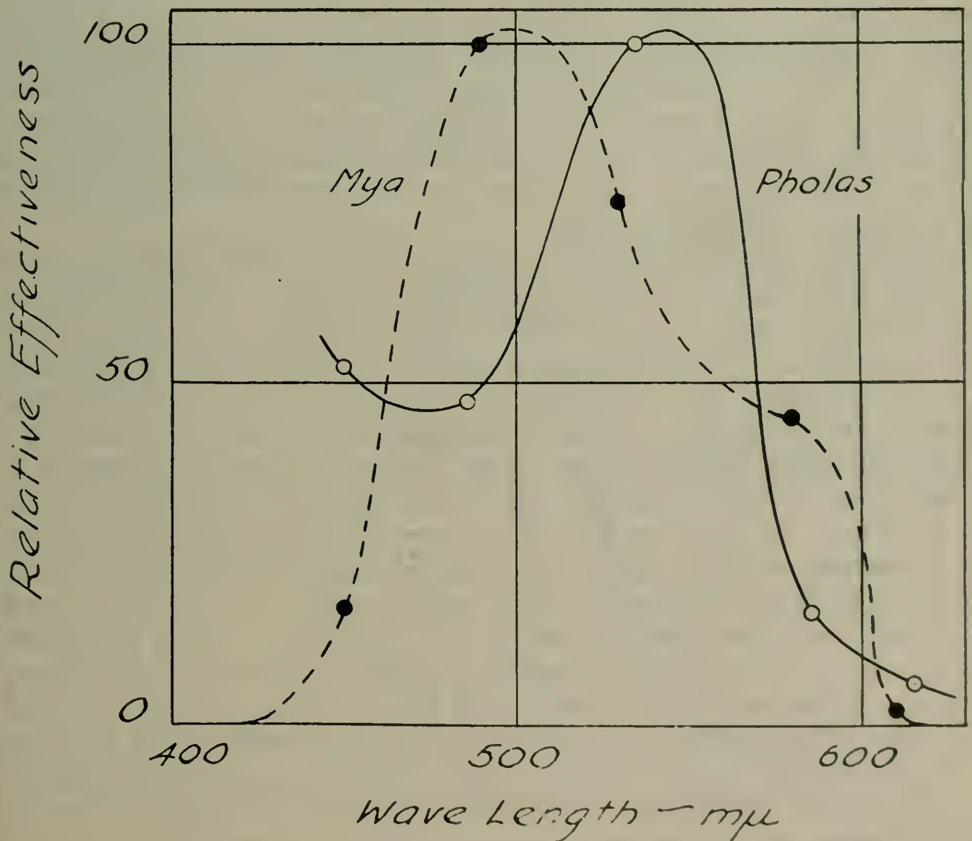


FIGURE 11

A Comparison of the Relative Effectiveness of the Spectrum in the Stimulation of Two Similar Mollusks, *Pholas dactylus* and *Mya arenaria* *dactylus,* which are both clams, and which possess a light sensitivity the kinetics and dynamics of which are extraordinarily alike. A study (29) of their sensitivity in the spectrum shows them to be quite different in this respect. The experiments are made by determining the energy required at different points in the spectrum to elicit the same sensory effect on the part of the animal. The reciprocal of this energy is proportional to the absorption coefficient of the sensitive system at different parts of the spectrum. Figure 11 shows these relationships for these two animals. It is apparent that the spectral sensibilities of the two are different, from which one may conclude that the substances S are different. This is merely negative evidence, since there has not been extracted from the sensitive areas of these animals a photosensitive substance which is known or identifiable spectrophotometrically as the one concerned. It has been possible to do precisely this in the vertebrate eye.

The rods of the vertebrate retina contain a substance which is sensitive to light. First discovered as a pigment by Boll (12), who called it visual purple, it was later studied in much detail by Kuehne (53), and still later by Koenig (45), by Koettgen and Abelsdorff (50), by Trendelenburg (77), by Garten (21), and more recently by myself (30, 31, 32). If visual purple is the photosensitive material (S) of the rods, then its absorption spectrum should be intimately related to the sensibility of the eye in the spectrum. Koenig (45) first made this comparison and found the visibility curve of the dark-adapted eye at low illuminations to coincide reasonably well with the absorption curve of visual purple. Later Trendelenburg (77) showed that the rate of bleaching of visual purple follows its absorption spectrum, and is also closely related to the visibility of the spectrum by the rods (Figure 12).

More precise work on the visibility curve of the spectrum for the rods (36) showed that this agreement is not quite accurate. Figure 13 shows that the visibility curve at low illuminations is about 7 or 8 $m\mu$ farther toward the red than is the absorption spectrum of visual purple, as measured in aqueous solution. This difference of about $7m\mu$ shows up even when the older data of Koenig and of Trendelenburg are carefully examined, as Figure 12 amply demonstrates. This need not, however, be taken as an invalidation of the conclusion that visual purple is the sensitive substance concerned in vision at low illumination. What the shift of $7m\mu$ shows is that the absorption spectrum of visual purple dissolved in the rods is slightly different from that of visual purple dissolved in water. Such shifts of absorption spectrum are known to occur very often in a pigment when it is dissolved in different media, especially if they vary in refractive index. This property has been called Kundt's rule (54) and, though having many exceptions, has been shown to be in agreement with a variety of facts. In its most general form it states that if one colorless solvent has a decidedly greater refracting or dispersing capacity than a second, then the absorption bands of a substance dissolved in the first will be nearer the red end of the spectrum than when dissolved in the second. The terminal segments of the rods are, from all observations and descriptions, fairly

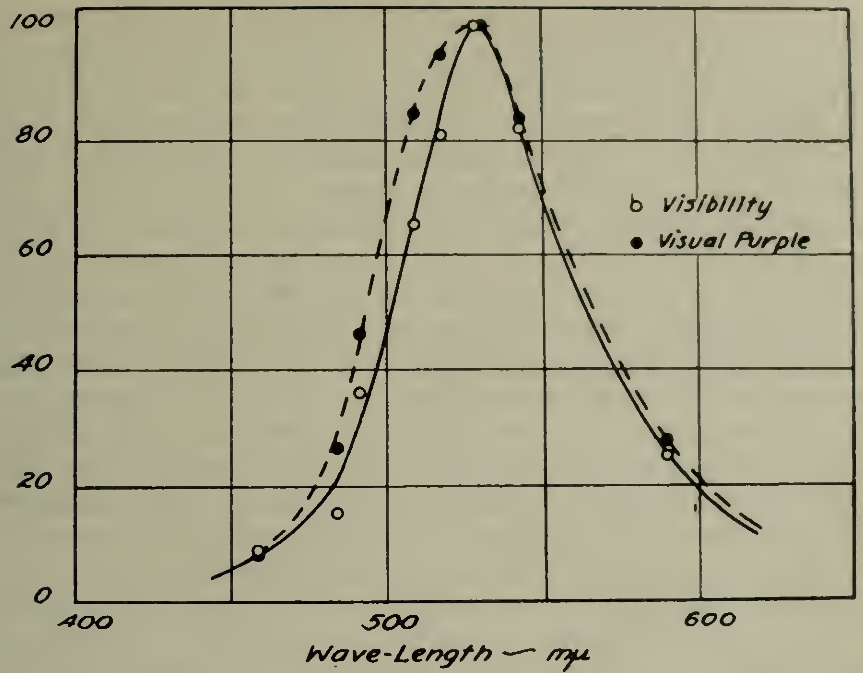


FIGURE 12

A COMPARISON OF THE RELATIVE EFFECTIVENESS OF THE SPECTRUM FOR THE VISION OF THE EYE AT LOW ILLUMINATIONS AND FOR THE PHOTOCHEMICAL BLEACHING OF VISUAL PURPLE

The data, taken from Trendelenburg (77), are only relative because they are not corrected for the energy distribution in the spectrum of the Nernst lamp used in the experiments.

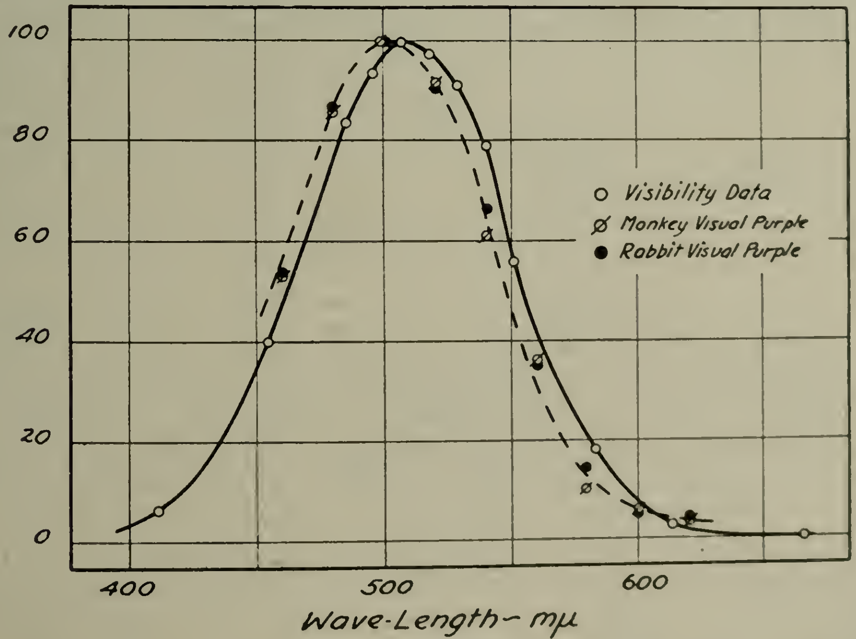


FIGURE 13

A COMPARISON OF THE EFFECTIVENESS OF THE SPECTRUM FOR VISION AT LOW ILLUMINATIONS WITH THE ABSORPTION SPECTRUM OF VISUAL PURPLE

The visual data are from Hecht and Williams (36) whereas the absorption spectrum data are from Koettgen and Abelsdorf (50).

dense and highly refractive bodies, and it is in them that the visual purple is dissolved in the living retina. It is therefore to be expected that its absorption spectrum in the rods will be farther toward the red than in watery solution. The small shift of 7mμ is thus in keeping with general knowledge and not only does not detract from but rather adds proof to the conclusion that the sensitive substance S present in the rods is visual purple.

From the agreement between the visibility curve and the absorption spectrum of visual purple, Koenig first drew the conclusion, which Trendelenburg also accepted as the result of his bleaching experiments, that in order to produce a visual sensation the incident light must be of such intensity that the visual purple in the rods will always absorb the same energy regardless of wave-length. It is in fact on the assumption that this type of conclusion is correct that the visibility curves of *Mya* and *Pholas* (Figure 11) may be interpreted as absorption spectra. It should however be added that this conclusion is true only as a first approximation. Modern quantum theory expects a slight but significant difference in the energy required for the two ends of the visible spectrum.

The photochemical properties which have been attributed to the substance S in the hypothetical photoreceptor system should be found in visual purple if the two are the same for the visual system of the rods. Studies in the photochemistry of visual purple have shown that this is true within the errors of such measurements. Visual purple in aqueous solution of bile salts is irreversibly bleached by light. The velocity of its bleaching should be proportional to the amount of light absorbed by the solution. For solutions either of small absorption or of small concentration this is equivalent to saying that the velocity of bleaching is proportional to the intensity of the incident light and to the concentration of the visual purple. If we rewrite equation [2], on page 219, somewhat more generally we get

$$\frac{dx}{dt} = k_1 I \; (a - x) \qquad [13]$$

which when integrated gives

$$k_1 I = \frac{1}{t} \log \frac{a}{a - x} \qquad [14]$$

which should describe the bleaching of visual purple. Here, as before, a is the initial concentration of visual purple, and x is the amount changed after t minutes' exposure to the intensity I. Equation [14] tells us two things which can be verified experimentally. One is that at a constant light intensity the bleaching reaction follows a monomolecular course. The other is that the rate of bleaching, as given by the constant $k_1 I$, is directly proportional to the intensity of the illumination.

The monomolecular course of the bleaching reaction is experimentally true (30) within the limits of error of the colorimetric method devised for this purpose. This consists of matching the color of a capillary tube containing about 0.1 cc. of visual purple solution against similar capillary tubes made of standard mixtures of bleached and unbleached visual purple. Table 4 gives an example of an experiment of this nature. It will be observed that the velocity constant k_1I may be considered constant, especially since the measurements are single determinations only, not averages, and are usually read only to the nearest five–per-cent concentration. Experiments specially designed to show whether the bleaching reaction shows any after-effect or latent, induction period failed to bring any such factors to light.

So far as the relation between intensity and velocity of bleaching is concerned, experiments (32) bear out the supposition that the velocity constant is directly proportional to the intensity of the incident light. Figure 14 gives a summary of four series of measurements made for the purpose of finding what this relation is.

It will be recalled that simple photochemical reactions show almost no change in velocity at different temperatures. Table 5 gives the results of two series of measurements (31) in which the photochemical bleaching of visual purple was carried on at three different temperatures. It is apparent that the velocity constant k_1I is uninfluenced by a difference of 30°C.

All this information renders it highly probable that so far as rod vision is concerned it has been possible to isolate and, to a certain extent, study independently the substance S. What this substance or substances may

TABLE 4

KINETICS OF DECOMPOSITION OF VISUAL PURPLE BY LIGHT

Time	Concentration observed	$k_1I = \frac{1}{t}\log\frac{a}{a-x}$	Concentration calculated
min.	per cent		per cent
0	100		100
6	65	0.031	66
12	42	0.031	44
18	25	0.033	29
24	20	0.029	19
30	15	0.027	13
36	10	0.028	8
46	0		
Average.		0.030	

TABLE 5

TEMPERATURE AND BLEACHING OF VISUAL PURPLE

Temperature°C.	$k_1I = \frac{1}{t}\log\frac{a}{a-x}$ Series 1	Series 2
5.2	0.039	0.031
20.0	0.035	0.031
36.1	0.036	0.033

be for the cones, one cannot say at present. I once thought (36), with Weigert (82) and with Hering, that it might also be visual purple in a more dilute form, and I explained the difference between the spectral position of the visibility curve of the cones and of the absorption spectrum of visual purple on the basis of Kundt's rule. But I now doubt whether such a hypothesis is fruitful, and am more inclined to believe that the sensitive substances in the cones are not visual purple, though they may well be closely related to it chemically. In a later section of this chapter I shall deal more fully with the problem of the number of individually different substances necessary for the cones.

VIII. Intensity Discrimination

The physicochemical system which I have suggested as a working hypothesis for the photoreceptor process applies to every cell in the sensitive surface. It, or a system like it, must form the ultimate basis on which the sensory process works. However, a sense-organ is not merely a group of isolated sense-cells connected with individual nerve-fibers. It is an organ, and functions as a whole. In this capacity it possesses certain

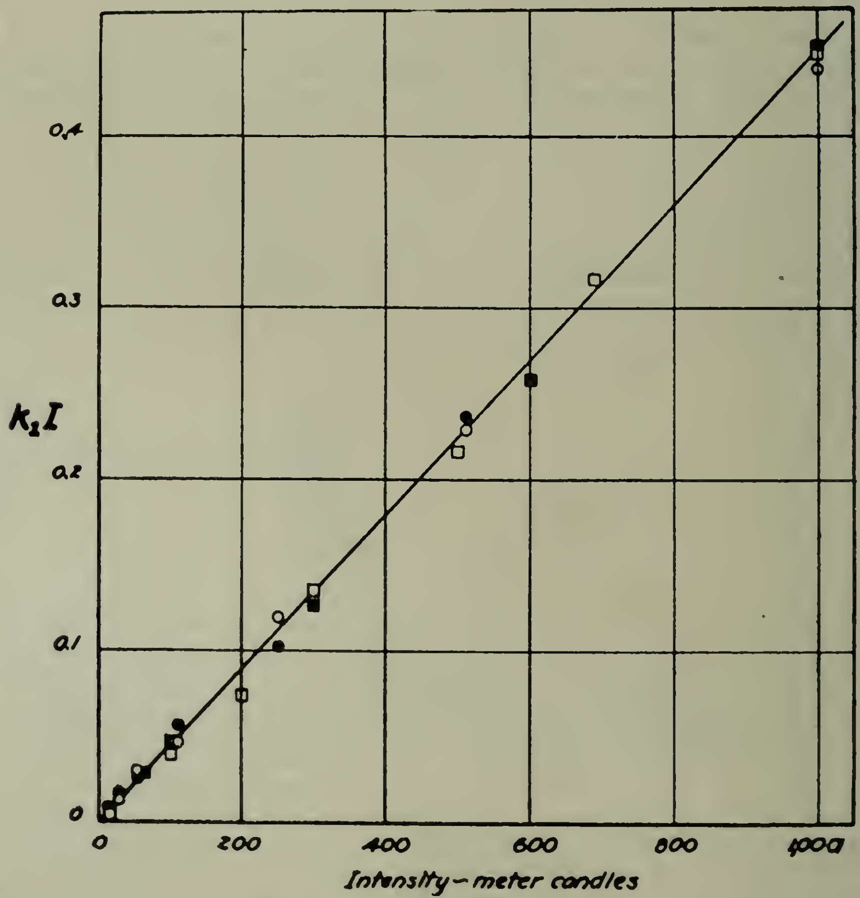


FIGURE 14

Relation between the Velocity Constant k_1I of the Bleaching Reaction of Visual Purple and the Intensity of the Light properties which depend primarily on the related presence of a group of sense-cells, and only secondarily on the particular mechanism by means of which each sense-cell acts. These properties of the sense-organ as a whole may be called physiological as opposed to the properties discussed up to now which may be termed physicochemical.

In order to understand the complete nature of the receptor process it is necessary to include these physiological characteristics and to find how they are brought about and what their basis is. I shall start with a property which is transitional, one that partakes of both physiological and physicochemical attributes. This concerns the matter of intensity recognition and discrimination. In our daily experience we constantly discriminate between bright and dim surfaces, and use this information in judging form and distance. Other animals evaluate intensity similarly. Intensity discrimination is thus a general problem. It has puzzled biologists for many years because of its importance in the relation between organism and environment, and between one part of an organism and another.

One phase of it has had a fascinating history culminating in what is currently known as the Weber-Fechner law. Weber (81) had found with the perception of weights that the minimum increase in weight which is just perceptible is roughly proportional to the weight already present. Calling ΔI the just-perceptible increase in the intensity I, he believed to have found that $\Delta I/I$ is constant. In a word, intensity perception is relative and not absolute. Fechner (19) later supposed that this ratio corresponds to a unit increase in sensation ΔS. He then integrated the relation between ΔS and $\Delta I/I$ and formulated the celebrated psychophysical law that the sensation produced by a stimulating agent is proportional to the logarithm of its intensity. This has been generalized to such an extent that when the physiological effect of a physical or chemical agent is proportional to the logarithm of its intensity, this relationship is referred to as the Weber-Fechner effect. For many years whenever the problem of intensity perception and discrimination was considered, it has been from the point of view of the Weber-Fechner law, and many ideas have been proposed to account for it in such terms.

The Weber-Fechner law, however, belongs to that class of generalization which, on account of its apparent reasonableness, becomes absorbed in the body of scientific knowledge and is held tenaciously as general truth, even though it represents only an extremely circumscribed portion of reality. Consider the curious sequence of events which has brought it to its present position. Weber in 1834 described $\Delta I/I$ as constant for weights; but it was in 1858 that Fechner thought he demonstrated its constancy for vision. It was almost immediately shown to be incorrect by Helmholtz (38) who found that $\Delta I/I$ is not constant. Within five years it was reinvestigated by Aubert (6) who showed conclusively that $\Delta I/I$ is not constant for vision, because it varies from 1/3 at low intensities to 1/146 at high intensities. But the Weber-Fechner law had been launched on its career and continued to be enthusiastically advanced by Fechner and others. However, the ratio

$\Delta I/I$ was again investigated by Koenig and Brodhun (48) some twenty years later in a magnificent piece of experimental research. In this they showed that over a specific but very small range of intensities $\Delta I/I$ may be considered constant; but that viewed as a continuous physiological phenomenon, $\Delta I/I$ decreased steadily as the intensity increased up to a certain point, beyond which $\Delta I/I$ increased. In Figure 15, I have brought together Aubert's experiments, the measurements of Koenig and Brodhun, and the recent work of Blanchard (9). It is apparent that though made at intervals of over half a century they give identical results with the human

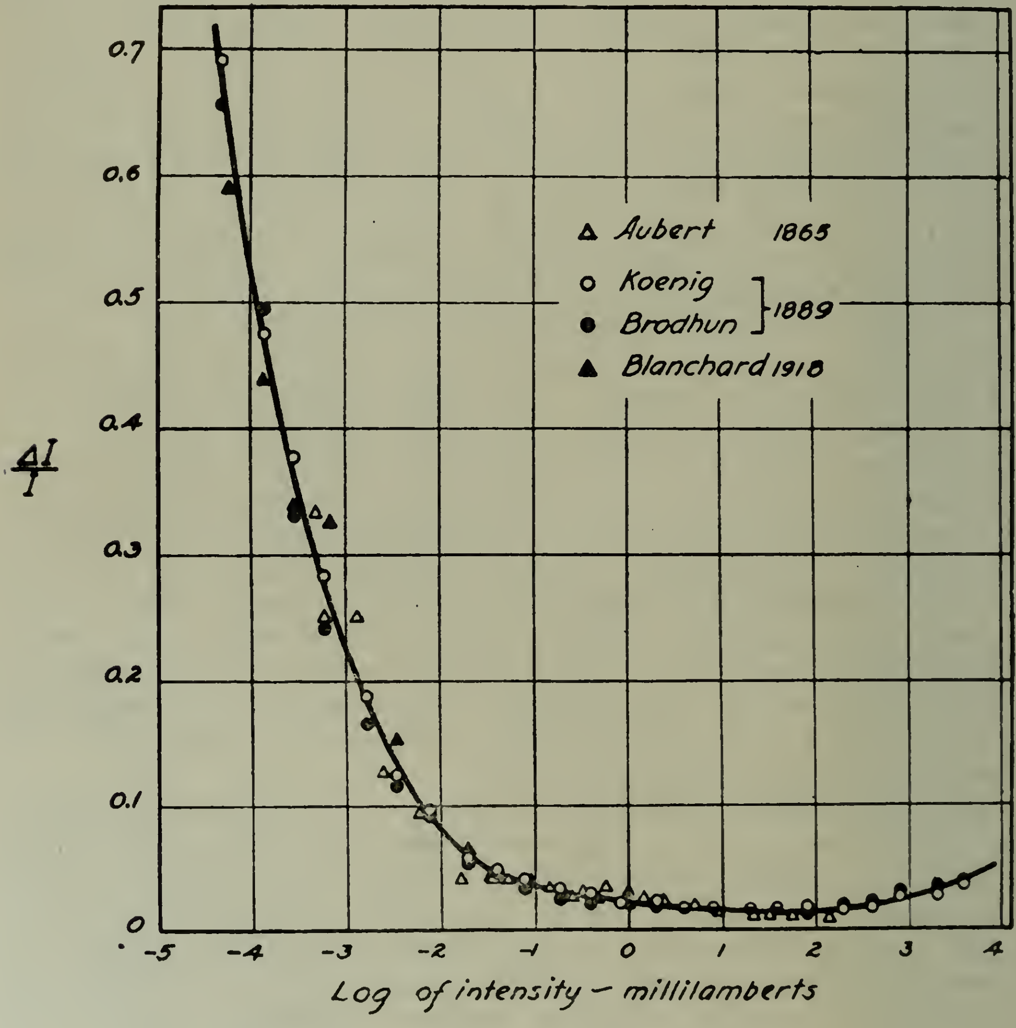


FIGURE 15

INTENSITY DISCRIMINATION OVER THE TOTAL ILLUMINATION RANGE VISIBLE TO THE EYE

Blanchard's data are given in their original units: millilamberts. Koenig and Brodhun's intensities have been divided by 250 to convert them into millilamberts. Similarly Aubert's intensities have been divided by 10,000. It is apparent that these three groups of investigators, working many years apart and independently, have arrived at essentially the same results.

eye. My own experiments (26) with intensity discrimination by the clam *Mya* show $\Delta I/I$ to have exactly the same type of behavior (Figure 16).

By an inversion, therefore, the problem of intensity discrimination can no longer be a matter of explaining why $\Delta I/I$ is constant, but rather why it varies as it does. As frequently before in this chapter, I shall begin an analysis of this problem with the comparatively simple system such as is found in the clam *Mya,* and proceed later to the more complex case of the human eye.

If an animal like the clam is exposed to light of intensity I, it responds by the characteristic retraction of its siphons at the end of a given time. If

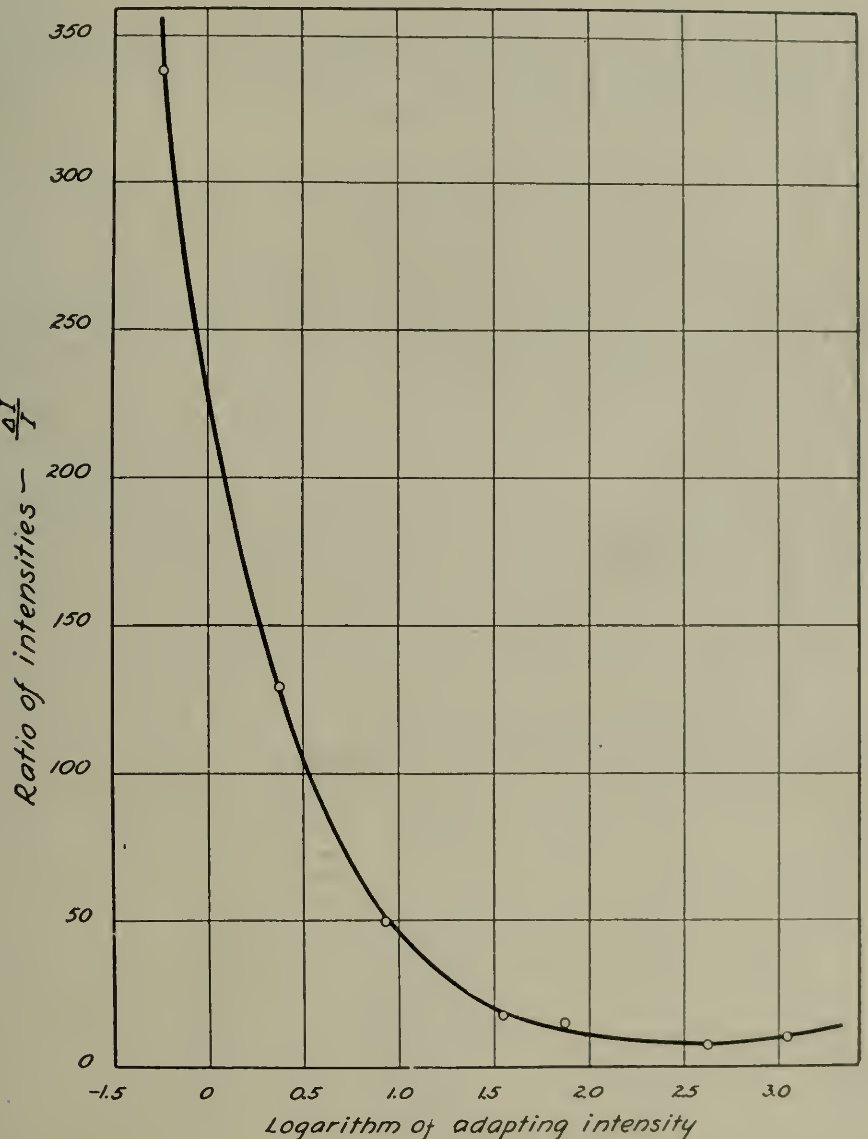


FIGURE 16

INTENSITY DISCRIMINATION IN *Mya*

Compare this curve with the data in Figure 15 for the human eye.

the illumination is maintained, the organism comes into sensory equilibrium with the light and nothing further happens. Again to elicit the usual response it is necessary now to increase the intensity of the light. The amount of the increase which is necessary to produce a response in a given time varies with the intensity to which the animal has already been adapted. The relation between the two is $\Delta I/I$, and has been measured over a large range of intensities. Table 6 and Figure 16 show the relation between $\Delta I/I$ and I for a response which occurs with a reaction-time of 2.1 seconds. This is not a minimal reaction-time, and so the values $\Delta I/I$ are very much larger than for the human eye. But experiments which measure $\Delta I/I$ for the just-perceptible response in *Mya* give values of the same order of magnitude as that for the eye. The way in which $\Delta I/I$ varies in Table 6 and Figure 16 is typical, and its analysis will therefore be generally applicable.

Let intensity I shine on the animal. The light reaction $S \rightarrow P + A$ will proceed, as we saw before, with a velocity

$$\frac{dx}{dt} = k_1 I (a - x_1) \qquad [15]$$

where $(a - x_1)$ is the concentration of S, and x_1 that of P and A. The reverse reaction, $S \leftarrow P + A$ will proceed as soon as some P and A are formed and will proceed according to

$$-\frac{dx}{dt} = k_2 x_1^2 \qquad [16]$$

If the light continues to shine, a stationary state will be reached in which these two opposing velocities will be equal. The equation

$$KI = \frac{x_1^2}{a - x_1} \qquad [17]$$

will describe this state, where $K = k_1/k_2$. If the illumination is now changed to $I + \Delta I$, then the stationary state

$$K (I + \Delta I) = \frac{x_2^2}{a - x_2} \qquad [18]$$

TABLE 6
INTENSITY DISCRIMINATION IN *Mya*

I meter candles	ΔI meter candles	$\Delta I/I$
0.574	194.1	338.2
2.38	307.	129.0
8.57	420.	49.0
35.4	612.	17.3
73.7	1,084.	14.7
412.5	2,355.	5.71
1,100.	10,230.	9.30

will describe the concentrations of S, P, and A. The photochemical effect produced by the addition of ΔI is therefore the increase in concentration $x_2 — x_1$.

The change in intensity from I to $I + \Delta I$ constitutes a stimulus for *Mya*. The resulting response is constant regardless of the experimental values of I and of ΔI. It is therefore very likely that the photochemical effect which produces the responses is constant. In other words, the change in concentration of S, that is $x_2 — x_1$, should be constant regardless of the curious way in which ΔI and I are related to each other. Table 7 gives the computations for the data given in Table 6. It is apparent that the values $x_2 — x_1$ are reasonably constant, especially in view of the large range in intensities and in values of x. This means that in order for an animal like *Mya* to distinguish between an intensity I and a specifically discriminable higher intensity $I + \Delta I$, the added intensity ΔI must result in a constant increment in the amount of photochemical change in the underlying photosensory system.

The human eye has two photosensory systems, the rods and the cones. In dark-adaptation these behave similarly but independently, as we saw. The same is true in intensity discrimination. I have computed (27) Koenig and Brodhun's measurements in terms of the same ideas which have just been used for *Mya*. Table 8 gives the data and the computations. It has been necessary to make corrections for the presumably variable pupil area and this has been done by means of Reeves's data (73). The values in the table show the following important points. At low illuminations intensity discrimination is determined by the rods. If in equations [17] and [18] for the stationary state we give K a value of 100, it is plain that the first seven intensities yield practically constant values of $x_2 — x_1$. The average value is 1.56, which represents the required photochemical change in the rods. For the next few intensities if calculated with $K = 100$, the quantity $x_2 — x_1$ drops sharply. If, instead, the value of K be put at 0.25, a new series of values for $x_2 — x_1$ are obtained smaller in magnitude

TABLE 7

PHOTOCHEMICAL CHANGES DURING INTENSITY DISCRIMINATION IN *Mya*

Initial intensity I	Conc. of $P + A$ x_1	Final intensity $I + \Delta I$	Conc. of $P + A$ x_2	Increase in conc. $x_2 - x_1$
m.c.	per cent	m.c.	per cent	per cent
0.574	2.4	194.7	35.4	33.0
2.38	4.8	309.4	42.3	37.5
8.57	8.8	428.6	47.5	38.7
35.4	17.1	647.4	54.3	37.2
73.7	23.7	1,158.	64.3	40.6
412.5	46.8	2,768.	78.0	31.2
1,100.	63.4	11,330.	92.5	29.1
Average				35.3

and roughly constant at an average value of 0.21. This is the required photochemical change in the cones.

Elsewhere, I have discussed the gradual drift in x_2—x_1 for the cones as shown in Table 8, and have indicated how a slight modification in treatment eliminates it. For the present purposes the constancy of x_2—x_1 is adequate. The values show, first, that x_2—x_1 is smaller for the cones than for the rods in a ratio of about 1 :8; second, that there is a sharp break at a certain illumination when intensity discrimination shifts from the rods to the cones; and, third, that x_2—x_1 for the rods and for the cones has in each case a constant order of magnitude. This constancy indicates that for the photosensory systems to discriminate between one intensity and the next perceptible one, the transition from one to the other must involve the decomposition of a constant amount of photosensitive materal.

We shall return to this conclusion in a moment. Meanwhile there are several interesting connections which these data and their analysis bear to those already presented, which it is important to point out. It will be remembered that $K=k_1/k_2$; therefore K is the ratio of the velocity constants of the light and "dark" reactions in the primary reaction of the photochemical system. Taking the obviously round values of K used in Table 8, we find that for the rods $k_2=0.01k_1$, whereas for the cones $k_2=4k_1$.

TABLE 8

PHOTOCHEMICAL CHANGES DURING INTENSITY DISCRIMINATION OF THE EYE

Illumination outside		Pupil area	Illumination at retina		Concentration			K
I	$I+\Delta I$		I	$I+\Delta I$	x_1	x_2	x_2-x_1	
0.0000484	0.0000800	50.3	0.00243	0.00402	4.81	6.14	1.33	100
0.0001336	0.000200	49.7	0.00664	0.00994	7.83	9.48	1.65	
0.000300	0.000400	49.5	0.0149	0.0198	11.48	13.12	1.64	
0.000644	0.000800	49.0	0.0316	0.0392	16.27	17.94	1.67	
0.001716	0.00200	48.6	0.0834	0.0972	25.01	26.69	1.68	
0.00358	0.00400	48.0	0.172	0.192	33.76	35.26	1.50	
0.00732	0.00800	47.7	0.349	0.382	44.15	45.59	1.44	
0.0190	0.0200	47.4	0.901	0.948	4.63	4.75	0.12	0.25
0.0382	0.0400	46.9	1.79	1.88	6.47	6.63	0.16	
0.0776	0.0800	46.3	3.59	3.70	9.04	9.17	0.13	
0.1948	0.2000	44.3	8.63	8.86	13.65	13.81	0.16	
0.3912	0.4000	42.0	16.43	16.80	18.32	18.50	0.18	
0.784	0.800	38.6	30.26	30.88	23.75	23.95	0.20	
1.956	2.000	33.6	65.72	67.20	33.14	33.45	0.31	
3.928	4.000	29.2	114.7	116.8	41.12	41.37	0.25	
7.86	8.00	24.4	191.8	195.2	49.30	49.59	0.29	
19.69	20.00	18.3	360.3	366.0	60.02	60.28	0.26	
39.36	40.00	14.2	558.9	568.0	67.45	67.71	0.26	
78.76	80.00	10.7	842.7	856.0	74.00	74.24	0.24	
196.2	200.0	7.4	1,451.9	1,480.0	81.64	81.88	0.24	
391.6	400.0	5.7	2,232.1	2,280.0	86.57	86.79	0.22	
776.8	800.0	4.5	3,495.6	3,600.0	90.61	90.83	0.22	
1,935.0	2,000.0	3.2	6,192.0	6,400.0	94.26	94.43	0.17	
3,853.0	4,000.0	2.2	8,476.6	8,800.0	95.68	95.83	0.15	

These do not tell us the absolute values of k_2 for the rods and cones because k_1 is probably different for the two. But they create a strong presumption that k_2 for the cones is larger than for the rods. The rate of dark-adaptation is a function of the "dark" reaction whose speed is determined among other things by the magnitude of k_2. From this it follows that the cones should have a more rapid rate of dark-adaptation than the rods. The data on this point are unequivocal. As Figure 10 shows, the dark-adaptation of the cones is practically complete in five minutes whereas the same process in the rods is not complete in 30 minutes.

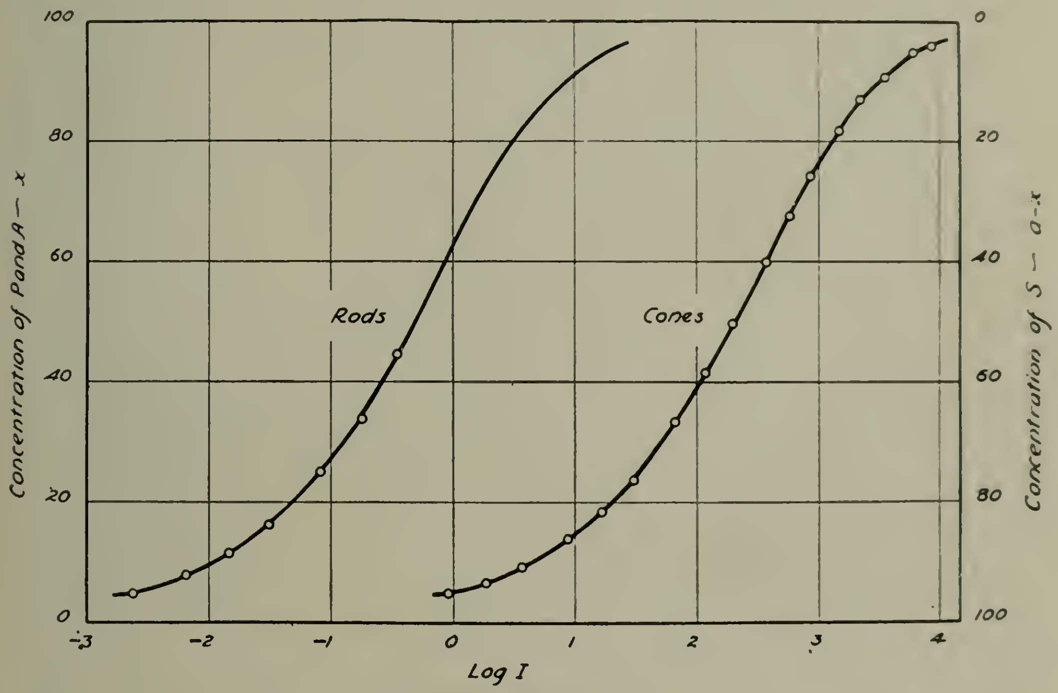


FIGURE 17

Relation between Illumination and the Concentrations of S, P, and A at the Stationary State of the Reaction $S \rightleftarrows P + A$ as given by the Equation $KI = x^2/(a - x)$

The abscissas represent retinal illumination; the ordinates are concentrations for the rods, when $K = 100$, and for the cones when $K = 0.25$.

An additional point of interest in regard to dark-adaptation is the relation between intensity and concentration of photosensitive substance or its precursors. In Figure 17 there is plotted the stationary state equation [17] for the rods and the cones using the values of K already given. The points are the x_1 values in Table 8. It is apparent that, between 10 and 90 per cent, both curves may without great distortion be treated as straight lines. Therefore, within a reasonable experimental error the concentration of sensitive material S may be considered proportional to —log I. It will be recalled that this is precisely the relation which we assumed on page 236 in order to interpret the data of visual dark-adaptation.

One further consideration may be added. The point of transition of intensity discrimination from rods to cones as computed in Table 8 comes at an intensity which from other independent measurements such as visual acuity is known to correspond to the transfer of function from the dim vision of the rods to the color vision of the cones. In Table 8 the last rod intensity is 0.0073 millilamberts. The transition point in the visual acuity measurements of Koenig (47) comes at the almost identical value of 0.0072 ml. There are many more correlations which may be shown between different aspects of vision. However, before going farther in this direction we must consider the matter of intensity discrimination from the point of view with which we began this section, that of the sense-organ as a whole.

Historically, the idea behind the Weber-Fechner law was an effort to describe the nature of intensity recognition in terms of intensity discrimination. It obviously failed to accomplish this because it was based on data which do not describe intensity discrimination correctly. However, the failure of the psychophysical law is deeper than this. It omitted recognition of a fundamentally significant characteristic of intensity perception, namely, its discontinuity. Starting with a given intensity, there is always a definite increase or decrease in this intensity which must be made before it can be recognized with certainty as different. In terms of our physicochemical explanation, each of these just-perceptible steps represents a constant increase in the quantity of photosensitive material decomposed in the visual cells. It is very important to consider the physiological nature and consequences of these step-like increases.

A naïve way would be to suppose that a greater concentration of decomposition products in the sense-cell represents a greater chemical or electrical effect on the nerve, which is then passed on to the proper centers and finally to an effector like a muscle or a gland. However, it has been known for a long time that nerve and muscle do not behave quite so simply; and recently Adrian and Zotterman (5) have shown that sense-organs do not respond in this way either. The impulse along a single nerve-fiber is constant in magnitude regardless of the size of the stimulus. The contraction of a single muscle-fiber is of constant magnitude regardless of stimulus. Nerve-impulses and muscle-contractions behave like quanta of energy in physics; they are either all or nothing. Just as the naïve conception of intensity has disappeared in physics and been replaced by number, so indeed is it disappearing in biology. A light that is more intense than another is merely emitting more quanta per unit time and space than another. Can we apply a similar type of idea in the perception of intensity in biological systems?

There are two ways in which such an idea may be considered, and both of them have been suggested. The oldest is that of Lucas (63) and was proposed for muscle. Lucas found experimentally that as the stimulus to a muscle is increased gradually in intensity the resulting height of contraction increases in a series of discrete steps, the number of which corresponds with the total number of individual muscle-fibers in the muscle. Intensity in muscle is therefore registered by the number of elements which are functional.

When it was shown (2) that the nerve-impulse is all or nothing as well, another idea was suggested by Forbes (20). This is that intensity is determined by the number of impulses which proceed over a single nerve-fiber in a given time. In other words, the frequency of the impulses determines intensity. In fact the researches of Adrian (3, 4, 5) have shown beyond a doubt that the frequency in a single nerve-fiber actually does increase with the strength of the applied stimulus, and can determine the height of contraction in an attached muscle.

Which of these two aspects of number applies to the intensity recognition of the photosensory system? Is it number of unit structures, or is it number of impulses in a single structure per unit time? One thing is certain. Whether we give it first place in vision or not, frequency is a factor in intensity perception. This is true, as Adrian's work has shown experimentally. The problem really is, first, whether number of structures enters at all as a factor; and if it does, second, how the concept of frequency of impulses in a single element and the concept of number of elements work together to give intensity discrimination and perception.

On the face of it, it is difficult to consider frequency alone as the basis of intensity discrimination. Frequency is a continuous function because time is continuous. But it is a fundamental property of intensity recognition that it is discontinuous in vision as well as in muscle. Therefore we must turn to the idea involving number of elements. This idea has long been in the air, having been suggested first perhaps by Brown (14), later by Forbes (20), and lately again by Lasareff (57). Its quantitative implications and demonstration have, however, not been made until recently (27, 34). The matter can actually be tested experimentally. Fortunately the experiments are at hand, and concern the relation between intensity of illumination and the visual acuity.

IX. Visual Acuity and Illumination

It is common experience that the fineness of detail which we can distinguish with our eyes varies with the intensity of illumination. The nature of this dependence of visual acuity on illumination has been known in a roughly quantitative way for about 150 years; but it was only 30 years ago that it was most carefully measured by Koenig (47) in such detail that his results, reproduced in Figure 18, have become classic. We know from these data that visual acuity varies with the logarithm of the intensity in a specific way.

This is an extraordinary situation when one tries to realize its implications. Visual acuity is, of course, the resolving power of the retina. The data, then, show that this resolving power gradually increases up to nearly 100 times its initial value as the illumination is increased. Now we know that the resolving power of a surface composed of discrete, independent, receiving elements—like the photographic plate for example—varies with the number of such elements in a given area, or, to be more specific, with their distance apart. The retina is such a surface, because it is composed of discrete elements, the rods and the cones. The data of visual acuity must therefore mean that at low illuminations there are fewer rods and cones present in the retina than at higher illuminations; certainly it must mean that the distance between the centers of the receiving elements varies inversely with the illumination.

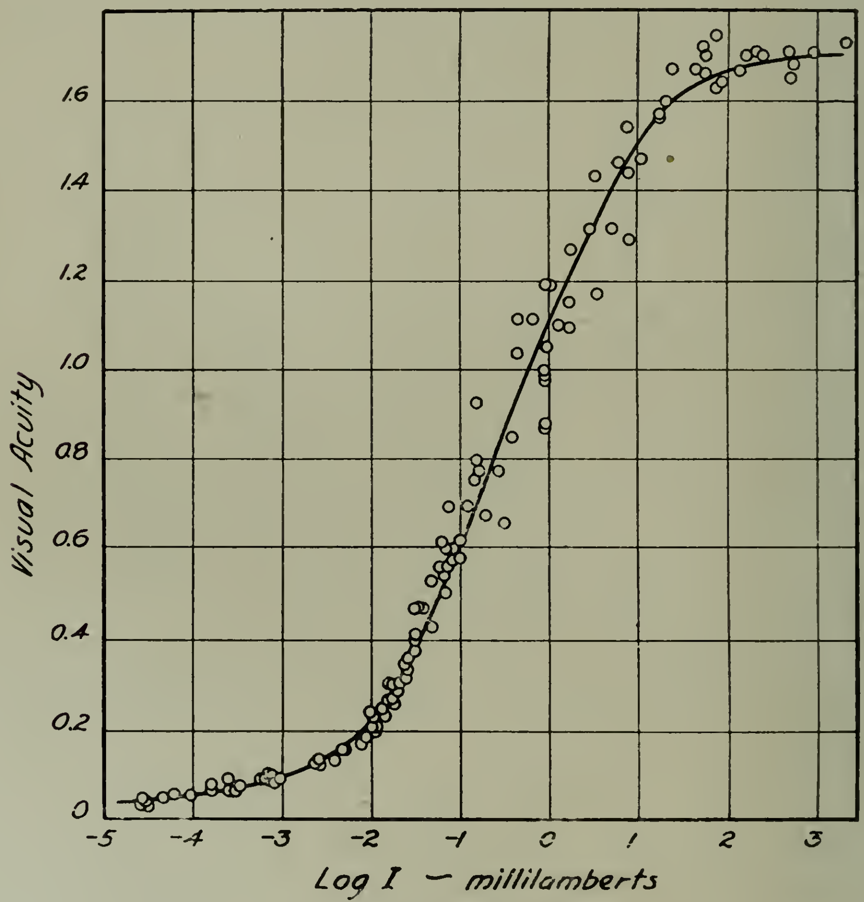


FIGURE 18

Relation between Visual Acuity and Illumination

Data of Koenig (47). The original intensities have been multiplied by 0.072 to convert them into millilamberts. The curve is a theoretical one derived as explained in the text.

But we know that this is nonsense anatomically, because the rods and cones are fixed in position in the retina; they do not separate and come together, nor do they increase and decrease in number. We must therefore assume these differences functionally. It is here that we realize that a sense-organ is not merely a single cell, nor a collection of isolated cells. The rods and cones represent two large populations of similarly constructed cells occurring within a given organ. We can suppose that they do not all have identically the same threshold, but that they vary in this respect much as other populations do. At low illuminations a few elements will be functional because the intensity is above their particular threshold, and as the intensity increases, more and more cells will become functional because the intensities will correspond to the thresholds of more and more elements. Since the thresholds are distributed at random, the distance between functional elements will decrease, and visual acuity will increase.

How can we describe quantitatively the relation between visual acuity and illumination in terms of such an idea? The simplest way is to ascertain what type of population distribution of thresholds it is necessary to assume in order to describe the data. We have seen previously that the rods and cones are two independently functional systems, which are to be treated separately. This is essentially the duplicity theory of von Kries (52). We may then draw two curves as in Figure 19 which represent the two populations in relation to their thresholds. It is apparent that

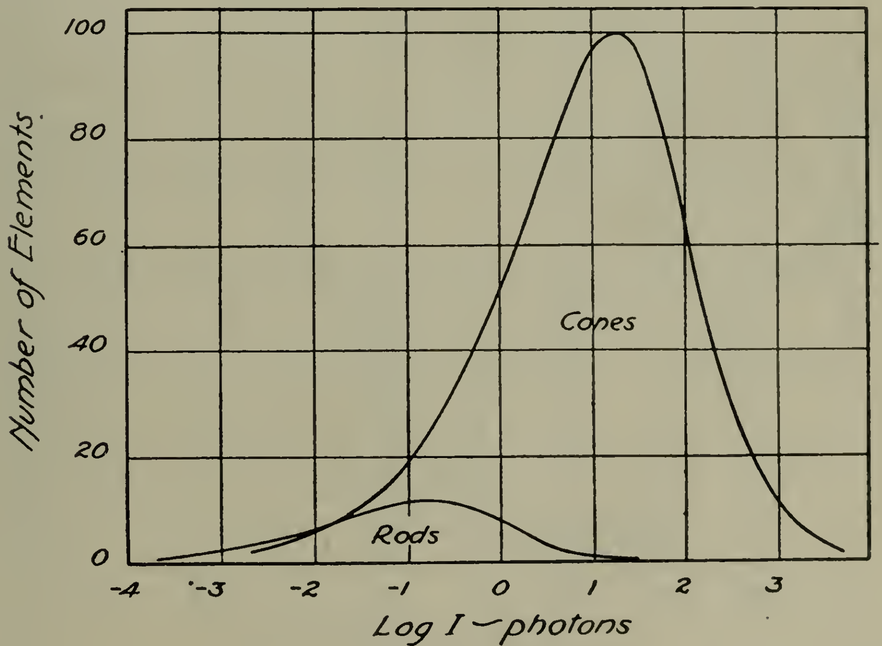


FIGURE 19

The Distribution of Thresholds of Rods and Cones Which has to be Assumed in Order to Describe the Data of Figure 18

The two curves are identical in form, but different in position and in size of ordinates.

the curves look like the common distribution curves used by statisticians. They are differential curves and give the number of elements whose specific thresholds correspond to a given intensity. What we need are integral curves which summate the individual thresholds and give us the total number of elements functional at a given intensity. These can be supplied by the usual methods, and are given in Figure 20.

These integral distribution curves show the situation clearly. At low illuminations a few rods are functional. They are therefore far apart and visual acuity is low. As the illumination increases the number of functional rods increases; they are closer together, and visual acuity increases. At a certain intensity, the cones begin to function. The number of active cones increases with the intensity, and, since the rate at which they increase

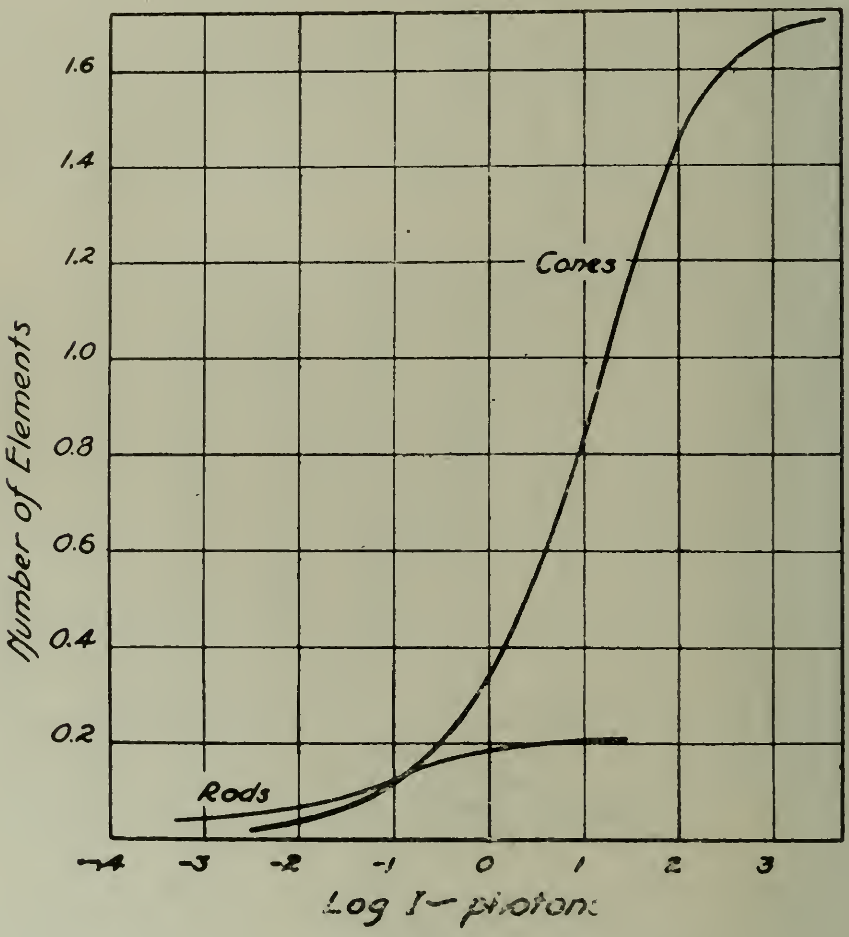


FIGURE 20

STATISTICAL DISTRIBUTION OF SENSIBILITY OF RODS AND CONES

These curves are the integrals of those in Figure 19 and give the relative number of elements functional at any intensity. The curves may be described by the common Gram series of the statisticians. However, the equation here used is $KI = x^2/(a - x)$ where K and a have different values for the two curves. The ordinates read directly in units of visual acuity.

is more rapid than for the rods, the functional cones soon become more numerous than the rods and therefore take over visual acuity. As they gradually increase in the number functioning, visual acuity continues to increase, until a point is reached when they are all functional and further increase in intensity produces no increase in visual acuity.

The curves in Figure 20 are corrected for pupil area by means of the data of Reeves (73). To show how accurately they describe the original data of Koenig the curve in Figure 18 has been drawn. This is nothing but the rod-cone curve of Figure 20 transferred back to the original uncorrected data by means of the pupil-area data of Reeves. Table 9 gives all the calculations for these data of Koenig, which for this purpose have been averaged in groups of five. It is apparent that the theoretical treatment is an adequate quantitative description of the relation between visual acuity and illumination.

There is one implication of this analysis which it is necessary to consider because of its historical and theoretical interest. The data heretofore given are for the normal eye, and the proposed explanation obviously rests on von Kries's (52) theory of the functional separateness of rods

TABLE 9

RELATION BETWEEN ILLUMINATION AND VISUAL ACUITY

Koenig's Data

Outside illumination I	Pupil area	Retinal illumination I	Visual acuity, x		
			Observed average	Calculated, rods $0.794I=\frac{(x-0.03)^2}{(0.21-x)}$	Calculated, cones $0.0906I=\frac{x^2}{1.71-x}$
0.0000297	49.6	0.00148	0.040	0.044	
0.000130	48.0	0.00625	0.068	0.058	
0.000390	46.6	0.0168	0.075	0.073	
0.000767	45.6	0.0350	0.093	0.087	
0.00269	43.5	0.117	0.13	0.12	0.13
0.00698	41.7	0.290	0.18		0.20
0.0112	40.3	0.451	0.22		0.24
0.0153	39.5	0.605	0.28		0.28
0.0214	38.5	0.826	0.31		0.32
0.0284	37.6	1.07	0.39		0.36
0.0397	36.5	1.45	0.45		0.41
0.0678	34.5	2.34	0.56		0.51
0.0826	33.8	2.79	0.61		0.54
0.137	31.9	4.36	0.72		0.65
0.215	29.8	6.68	0.75		0.76
0.585	25.4	14.9	1.00		0.99
0.980	22.7	22.2	1.00		1.10
1.47	20.7	30.4	1.12		1.19
3.57	16.5	59.0	1.30		1.38
8.93	12.9	115.0	1.44		1.49
26.6	9.3	248.0	1.61		1.62
65.9	7.1	468.0	1.69		1.65
155.0	5.4	838.0	1.68		1.66
1096.0	3.1	3400.0	1.71		1.70

and cones. Koenig recognized this when he attributed the lower limb of the data in Figure 18 to the rods, and the remainder to the cones. He also recognized the implications involved, namely, that in a completely color-blind eye the cone portion of the curve should disappear and leave only the lower rod limb plus any extensions of it. Figure 21 gives the data for two such individuals, one measured by Uhthoff (80) and the other by Koenig (47). These bear out the supposition that the two limbs of the normal curve follow von Kries's duplicity theory.

These data are equally significant for our explanation. Figure 20 shows that for the normal eye the cones overtake the rods at the middle of the rod range. This it will be recalled from Figure 16 is true also for intensity discrimination. It therefore follows that in a completely color-blind eye the visual acuity data should extend to a distance beyond the rod-cone intersection point equal to that which has preceded it. Moreover, the entire visual acuity data of such a case should be described by only the rod curve of Figure 20.

The lines drawn in Figure 21 are this rod curve taken from the theoretical curve of Figure 20 and corrected for pupil area as already explained. It is apparent that both suppositions are correct. The data of the completely color-blind extend over twice the range of the rod vision of the normal. There is no rod-cone inflection point, and the entire data are describable in terms of the rod distribution curve alone.

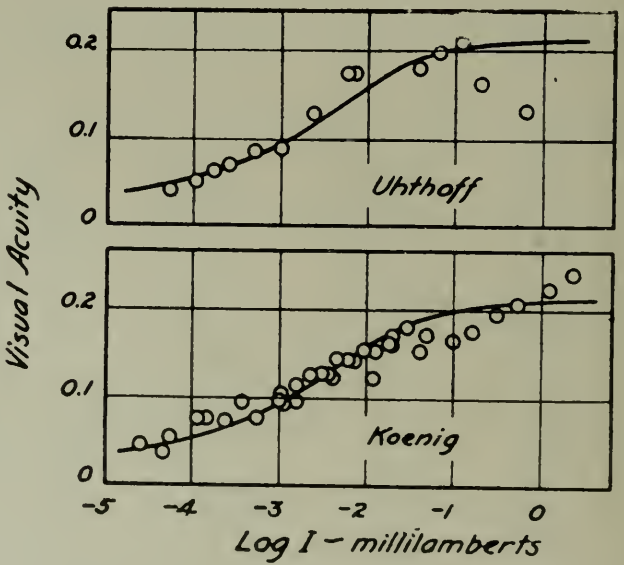


FIGURE 21

Visual Acuity of Two Completely Color-Blind Individuals

The data of Koenig (47) have been multiplied by 0.072, and those of Uhthoff (80) by 0.000169 to convert them into millilamberts. In addition Uhthoff's units of visual acuity have been multiplied by 0.75 to render them comparable to Koenig's. Compare this figure with Figure 18, but note that the ordinates here are twice as large as in Figure 18.

In the complementary case of a night-blind person the rod portion of the curve should be lacking, and the entire visual acuity curve should be describable by the cone curve alone. Such a case has not been studied in this way, but it is confidently expected that this will be the result.

These experiments and their theoretical treatment have been confined to the human eye. In order to determine whether the ideas involved have any general validity in the physiology of the visual process, Wolf and I (37) studied the relation between visual acuity and illumination in so differently constructed an organ as the eye of the bee. To do so we had to develop a method for the investigation of the vision of animals other than man. Starting with the common observation that animals with eyes respond to a sudden movement in their visual field, we converted it into a precise method of measuring visual acuity.

Our measurements, made with worker bees, are given in Figure 22. It is at once apparent that the visual acuity in the bee varies with the illumination in much the same way as in the human eye. Figure 23 represents the differential, population distribution of thresholds among the ocular elements which has to be assumed in order to describe the data in Figure 22. It is, therefore, apparent that essentially the same theoretical basis that we found for the human eye underlies the visual acuity of the bee's eye as well.

X. INTENSITY, NUMBER, AND FREQUENCY

The relation of all this study of visual acuity to the problem of intensity perception and discrimination is almost obvious. Here we have completely independent evidence of the fact that differences in intensity are associated with differences in the number of elements functional in the retina as a whole. However, for us to link intensity discrimination with this variation in number we need a closer consideration of visual acuity and intensity discrimination. On the basis of the same intensity discrimination data which we have used in Section VIII, Koenig (46) found that it is possible to recognize 572 discrete steps in intensity over the complete range visible to the eye. To judge by our analysis of these data, the first 30 of these steps are mediated by the rods; the remainder, 542, by the cones.

If the recognition of a minimal intensity difference ΔI represents a change from n to $n+1$ or $n-1$ in the number of cones functional, then there must be at least 542 cones in the fovea which are concerned in the process. These cones cannot be conceived as lying any place on the fovea. They must be thought of as located in a functionally related area—a unit retinal area. This is a new concept in the physiology of the receptor process, and no longer concerns the sense-cells as individuals, but rather as related entities. This area must have 542 cones because only in this way can each step in intensity discrimination correspond to an increase or a decrease in the number of functional cones in a unit area.

If visual acuity and intensity discrimination are manifestations of the same mechanism, such a minimal retinal area should also be able to mediate all the visual acuities, from the lowest to the highest. This area will have

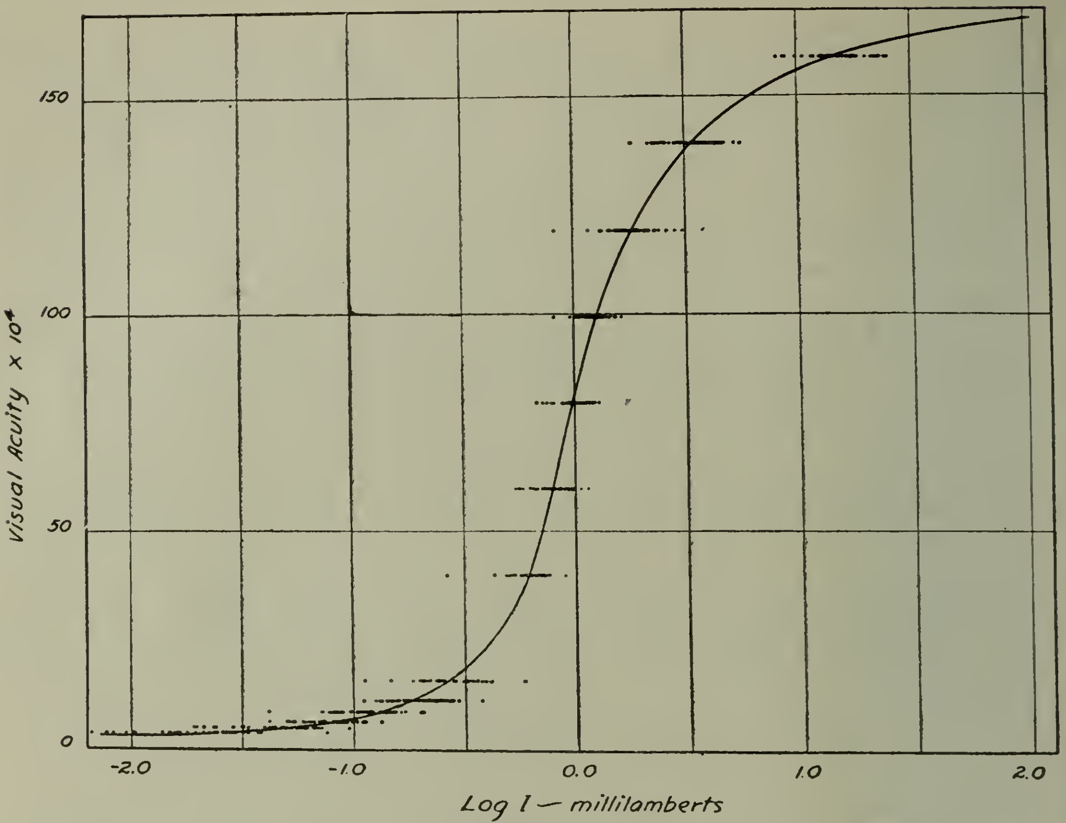


FIGURE 22

Relation between Illumination and Visual Acuity of the Eye of the Honeybee

Each measurement with each animal is represented by a dot. It is apparent that visual acuity varies with log *I* very much as it does in the human eye.

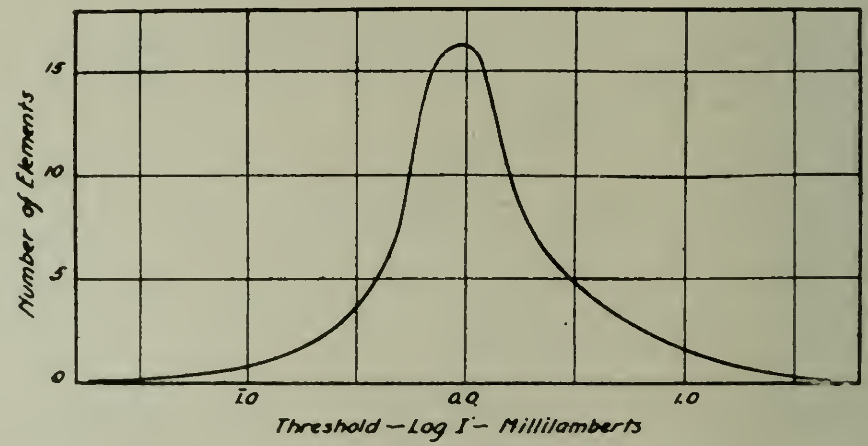


FIGURE 23

Statistical Distribution of Thresholds of the Ocular Elements (Ommatidia) Which Has to be Assumed for the Bee's Eye in Order to Describe the Data of Visual Acuity in Figure 22 for its linear dimensions the retinal distance between two just-perceptible contours corresponding to the lowest visual acuity. The retinal distances for all other visual acuities will obviously fall within this distance. From Koenig's data the lowest visual acuity is 0.03 units. This corresponds to a visual angle of 44′, and to a retinal distance of 0.2 mm. The resulting square area is 0.04 sq. mm. The fovea contains 13,500 cones per sq. mm. This minimal visual acuity area therefore contains 540 cones, which is the same number as that derived from measurements of intensity discrimination.

Another relation between visual acuity and intensity discrimination can be shown as follows. The first 30 steps in intensity discrimination are made by the rods. It was shown that this corresponds to just about half the full range of function of the rods. The same is apparent when the rods give up the determination of visual acuity to the cones. The full possible range of rod activity in intensity discrimination is thus 60 steps. The cones have 542 steps. This is a ratio of 1 : 9. We have already found the same ratio for the photochemical effect required for each step. Note, moreover, the same ratio in visual acuities. The maximum rod visual acuity is 0.21; the maximum cone visual acuity is 1.71. This is a ratio of 1 : 8+.

It can be shown even more directly that visual acuity and intensity discrimination are two manifestations of the same mechanism when we consider the basis for the population curves of Figure 20 describing visual acuity and illumination. We have called these statistical population curves; and so they are, because they look like such curves and can be described by the usual statistical equations. However, they are also the curves which describe the concentration x of the photochemical decomposition products P and A at the stationary state produced by an intensity of illumination. The integral curves are the graphs for the stationary state equation [17]

$$KI = \frac{x^2}{a - x}$$

which we used for intensity discrimination, and this equation has actually been used in the computations in Table 9. The relation between the two rests on the general supposition that each concentration of photochemical products has the same number of cells whose discharge threshold it represents (27, 34). Thus for number of cells in the integral population curves one may substitute concentration of $P+A$ and the curves become the stationary state curves of intensity discrimination.

The same photochemical mechanism therefore underlies both properties, and we must conclude that intensity recognition as well as visual acuity is determined by the number of retinal elements functional in a given retinal area. What then becomes of the frequency of discharge of a given cell? This surely exists in the retina as Adrian has shown, and undoubtedly has a function. We must therefore consider its significance if the present treatment of intensity discrimination is to be adequate.

I have emphasized the fact that intensity recognition is discontinuous. With this discontinuity is, however, associated a curious property. Let us place intensity I on one side of a photometric field, and the next just-perceptible intensity $I+\Delta I$ on the other side of the field. According to our ideas, I corresponds to n cones per unit area and $I+\Delta I$ to $n+1$ cones. What happens if we were to put an intensity midway between these two, say $I+\Delta I/2$ on one side of the field? How many cones would this correspond to? Obviously to n again. Now what is the next just-perceptible intensity to this? Experiment shows that it certainly is not $I+\Delta I$ which presumably corresponds to $n+1$ cones. The next perceptibly different intensity is very nearly $I+\Delta 1/2+\Delta I$. This seems absurd since if $I+\Delta I/2$ corresponds to n cones, then all that one should need to go to is $I+\Delta I$ previously known to correspond to $n+1$ cones. In other words, though intensity discrimination is discontinuous, there are no critical points in intensity recognition.

The solution of this paradox lies in a recognition of the place of frequency in the photoreceptor process. Consider any sensory cell under continuous illumination. Light energy enters at the receiving end. It is transformed in the body of the cell into photolytic products, which by their proper reactions result in an impulse coming out of the end connected with a nerve-fiber. The actual amount of photochemical change required for the cell to discharge an impulse varies with the position of the cell in the population curve of Figure 20 as I have already indicated. Since the illumination is continuous, there will be a continuous formation of photoproducts which will accumulate and discharge rhythmically at a rate proportional directly to the concentration of $P+A$ (or to the rate of formation of B) and inversely to the discharge threshold of the particular cell. As the intensity of the light is increased, the frequency of this discharge will increase steadily. Presently the threshold of the next most sensitive cell will be reached, and it will begin to function at a given frequency, which will also increase with intensity. Since we found intensity discrimination to correspond to a definite increase in photochemical products, it must represent the bringing-in of an additional element which is functioning at the same frequency as the last one.

This means that intensity discrimination depends on a definite increase in the total frequency of discharge of a series of related retinal elements, the increase being produced by the addition of a new functional element. In this way there may be brought together under one general set of ideas not only visual acuity, intensity discrimination, and the paradoxical discontinuity of intensity recognition without the presence of critical intensities but the all-or-nothing law, the known increase in frequency of sensory impulses with intensity, and the relation between number of elements functional and the activity of an organ like the eye or a muscle.

XI. Color Vision

The two major factors with which the photosensory process is concerned are intensity and color. I have devoted considerable space to an analysis of intensity perception and its related phenomena because it has been possible to present it from the unified point of view developed in terms of the physical chemistry and physiology of the receptor process. I wish that it were possible to do the same for color vision. There are many well-established facts of color vision, some of which have been measured with considerable accuracy (79). But there is no theory which has described these diverse facts quantitatively in terms of a set of assumptions representing a mechanism of color vision. There are, however, one or two things about color vision which are so fundamental that they need to be related to what we have already learned of the photoreceptor process.

Color vision is the capacity of the eye to divide the visible spectrum into a series of regions which produce qualitatively different sensory effects. The normal human eye can separate the visible spectrum with complete certainty into about 180 patches which cannot be made to look like one another by varying their intensity. The classical researches of workers from Newton to the present have shown that these 180 patches, in fact that all the known hues including those outside the spectrum, can be duplicated by, and therefore described in terms of, three variables made up each of a monochromatic part of the spectrum.

This is fact, and has not changed in a hundred years. The particular specifications of these three variables may vary, and indeed may be quite arbitrarily assumed. Nevertheless, they are adequate as a formal description of the properties of color vision (79). Such a basic fact should be utilized and incorporated into any mechanism proposed as a basis for color vision. This indeed was done about 125 years ago by Thomas Young (83). Young supposed that there are three kinds of "fibers" in the retina, each producing a characteristic sensation, one of red, another of green, and a third of blue (violet). Each type of fiber is sensitive practically to the entire visible spectrum, but the first possesses a maximum of sensibility in the red, the second in the green, and the third in the blue. The various color sensations then result from the relative strengths with which the three different fibers are stimulated by the objective light.

Few people today suppose that Young's idea, as here given in its simple form or as elaborated by Helmholtz (39), Koenig (49), von Kries (52), and others, is adequate as a complete theory for the mechanism of color vision. The question, however, has often been raised as to whether it can serve even as the basis on which to construct a theory of color vision.

The main objection to Young's hypothesis is founded on the uniqueness of the sensations of yellow and of white. Mixtures of green and blue lights give a continuous series of colors in which both are identifiable. The same is true with mixtures of red and blue lights. But in mixtures of red and green a new sensation, yellow, arises which contains neither red nor green. The same is true for mixtures of yellow and blue, or of green, red, and blue, which give white, a sensation *sui generis*. If Young's idea is correct then it must be supposed that yellow is a central phenomenon which occurs when in the retina red and green receiving fibers function simultaneously. Similarly, white occurs in the brain when all three—red, green, and blue—

receiving fibers function in the retina. It is precisely against such a formulation that criticism has been raised by Hering (41) and by Ladd-Franklin (55). And as a result, first Hering and later Ladd-Franklin have devised theories which assume separate receptor substances for yellow and for white.

If there is to be developed an adequate theory for the mechanism of color vision, a decision must be made in the very beginning as to which of these two conceptions of yellow and white is correct. Are there special substances in the retina for the reception of yellow and white, or are yellow and white phenomena which arise in the brain out of the impulses coming from the three kinds of fibers (or substances) in the retina? The answers to these questions can be given experimentally.

If red light and green light fall on a given retinal area of one eye and a yellow sensation results, it is not possible to decide among the following interpretations of what happens: (*a*) the photochemical product of the action of red light combines chemically with that formed by green light to give a "yellow" substance which stimulates the proper nerve-endings, as Ladd-Franklin postulates; or (*b*) a substance reversibly sensitive to red and green remains unchanged while a substance reversibly sensitive to yellow and blue is changed by the red light to which it is partially sensitive, as Hering proposed; or (*c*) the red sensitive fibers and the green sensitive fibers are both active and the brain synthesizes yellow, as must be the case if Young's idea is adopted. But if red light falls on a part of the retina of one eye, and green light falls on the corresponding portion of the retina of the other eye and the result is a yellow sensation, then (*a*) and (*b*) cannot be true, and only Young's idea is tenable. This is precisely what happens in the binocular mixing of colors.

The problem of the binocular mixing of colors is an old one; it had been a subject of controversy before Helmholtz. But it was Helmholtz in his *Physiologische Optik* who settled the matter temporarily by stating that he was unable to fuse colors binocularly. This has been generally accepted even though Hering (40) later showed the cause of Helmholtz' failure and arranged an apparatus to overcome it; and von Bezold and Dobrowolsky (17) defined quite clearly the difficulties involved and avoided them. As a result of the further suggestions by various people, one finds von Kries in the 1911 edition of Helmholtz' *Physiologische Optik* (39) admitting the possibility of binocular color fusion, though he finds it difficult to accomplish.

Since then Trendelenburg (78) has constructed a binocular modification of the classical Helmholtz monocular color-mixing apparatus and has demonstrated beautifully that not only can binocular color fusion take place but that one can obtain regular color-mixing equations comparable to those obtained monocularly. Red and green binocularly give yellow; and several complementary pairs give white. However, even with this improved apparatus the process is described as difficult and requiring considerable practice. Recently Rochat (75) has devised another type of spectroscopic apparatus for binocular color mixing and obtained the same results as Trendelenburg. Here again emphasis is laid on the difficulties involved and the practice required.

It must, therefore, be concluded that modern work has definitely settled the problem of binocular color mixing. Red in one eye and green in the other on identical retinal areas result in a yellow sensation. This yellow obviously cannot arise anywhere but in the central nervous system. The only difficulty seems to be that binocular color mixing is reputed to be a hard feat to perform and to require much practice. Even this, however, is no longer true, because of a very simple arrangement of colored filters, by means of which it is possible to secure binocular fusion easily and without any previous practice (35). This experiment is of such critical importance, and its point has been so completely missed up to the present that I give briefly a method of demonstrating it which should be found in every interested laboratory.

Use is made of two color filters—Wratten filters procurable from the Eastman Kodak Company—No. 29 (red) and No. 58 (green) because they transmit light of about the same brightness. A blackened box 30 cm. long, 15 cm. wide, and 10 cm. high is arranged so that one 15 x 10 end is open and fits against the face. The other 15 x 10 end is closed, and contains the two 5 cm. square Wratten filters, one next to the other, and in a light-tight fit. About a meter from one's eye is a white cardboard with a bright light as near in front of it as possible. We use a 250-watt or a 100-watt concentrated-filament lamp projecting through a hole in the cardboard. This illuminates the cardboard and serves as a point of almost forced fixation for the two eyes. On looking at the bright light with both eyes focussed on it through the box, one immediately sees a bright yellow light in the center of a yellow square of cardboard flanked on one side by a green square and on the other by a red square.

What actually happens one can determine by looking with one eye at a time. One eye looks at the light through the green filter, and at a neighboring piece of cardboard through the red filter. The other eye looks at the light through the red filter and at a neighboring piece of cardboard on the other side through the green filter. By focussing both eyes on the light the red image in one eye and the green image in the other eye fall on identical points of the two retinas and are fused in the brain to give the appearance of a yellow light. The neighboring green in one eye and red in the other cannot be made to overlap at the same time and, therefore, are seen monocularly as green and red. The total sensation is somewhat as if one were looking at the light and the cardboard through three colored openings, a red (monocular), a yellow (binocular), and a green (monocular). The effect appears instantaneously if the eyes are focussed correctly. No explanation, except the injunction that both eyes look at the light, is necessary for performing this binocular fusion.

The green and red Wratten filters in the box may be replaced by yellow and blue filters to make binocular white. Wratten filters 16 (yellow) and 44A (blue) are a good combination for this purpose.

The binocular formation of yellow and of white, therefore, shows that such theories as those of Hering and of Ladd-Franklin, which require special receptors for yellow and white, are neither necessary nor correct. The uniqueness of yellow and of white as sensations is no obstacle to adopting Thomas Young's three-receptor idea as a basis on which to build a theory for the mechanism of color vision.

This is only the very first step. In terms of the photoreceptor mechanism which we have been considering, it means most probably that there are three different types of cones, each type having a different substance S as its photosensitive material. The three substances need not be very different chemically one from the other; they may all be only slight modifications of a fundamental photosensitive material.

It would seem that the next step in the theoretical analysis of color vision is a quantitative description of the spectral properties of these three types of cones and of their sensitive substances. Efforts to define them have not been lacking. Since the first attempt of Maxwell and of Helmholtz (38), the work of Koenig and Dieterici (49), followed by that of Ives (43), Exner (18), Abney (1), and others have given us three curves, currently known as "the three sensation curves" (79), which are supposed to do precisely this. These curves are derived from the measurements of color mixing. But since their only function is to describe the data of color mixing, they hardly represent progress in the theory of color vision. They cannot be used to describe any other properties of color vision except in the roughest possible manner, and inadequately at that. Moreover, there is nothing particular about their shapes or positions, since by linear and homogeneous transformations a great variety of three curves may be derived from them which will perform the same function of reproducing the data of color mixture (44). The next step after the adoption of the trireceptor idea—that is, the quantitative description of the three receptors—still remains to be done.

XII. General Considerations

This chapter began with a discussion of the similarities of different sense-organs. I pointed out that the very first part of the receptor process must be different for each kind of receptor; that there is a specific affinity between the sensitive substance and the particular part of the environment with which it is concerned. The final sensory effect in the central nervous system is also different for each sense-organ. It might then seem as if the action of the environment on the specialized substance is the thing which produces the specificity or quality of the resulting sensory response. We know, however, that such a simple and direct relation between outside reality and its sensory effects does not exist. Whatever the external agent may be which initiates the activity of a sense-organ, the final thing which the receptor organ does is to impart an impulse to the afferent nerve attached to it. And nerve-impulses are identical. How then is the specific quality of the environment transferred to the brain?

This situation was fully appreciated a hundred years ago by Johannes Müller (65). Müller pointed out, first, that different receptors produce qualitatively specific sensory results and, second, that no matter how a given receptor is brought into play, its sensory effect is qualitatively constant. A visual response is always associated with stimulation of the eye whether this stimulation is produced by light, by pressure, or by the electric current. His conclusion was that the particular stimulus used has nothing to do with the specific activity of a given sense-organ. Instead he supposed that each sense-organ complex possesses a definite "energy" or "quality" which produces its specific sensory effect—the eye, vision; the ear, hearing; and so on—and this quality is called out no matter how the sense-organ is brought into function. Müller thought that this specific energy is associated with the nerve which connects the sense-organ with the brain. But he fully realized that the specificity may not reside in the afferent nerve but might be even deeper in the central nervous system.

Modern work has served to corroborate and amplify Müller's analysis. We now know that nerve-impulses, no matter what their origin, are essentially the same. Motor impulses are like sensory impulses; excitatory impulses are like inhibitory impulses. The function of a particular nerve-fiber depends on its termination. For example, by the proper crossing and subsequent union of such specifically different nerves as the cervical sympathetic and the vagus, Langley (55*a*) showed that impulses which normally produce cardiac inhibition can be made to produce contraction of the arterioles of the ear, and vice versa. The specific effect produced by a nerve-impulse is determined by what type of tissue it ends in, and by the kind of ending it has in this tissue. In other words, qualitative differences in the effect of nerve-impulses are accomplished by structural prearrangement.

This is very significant for an understanding of the different sensory qualities shown by different sense organs. It is equally significant for a proper comprehension of the function of a single sense-organ like the eye. Not only is each type of receptor specialized in possessing a specifically susceptible material for normal contact with a particular part of the environment, but this specialization extends even within a given sense-organ. In the eye we have been forced to assume first a division between rods and cones (duplicity theory) and later a threefold subdivision among the cones (Young's trireceptor theory). Each of these four kinds of photoreceptor is concerned with a different aspect of the light environment—the rods with low illuminations, and the three cones with three parts of the visible spectrum. Moreover, each of these four subdivisions of the organ of vision has its own, qualitatively different, sensory expression. In other words, we must apply Müller's doctrine of specific energies to the four separate sensory qualities with which the visual receptor process is concerned. Since this specificity is not in the impulses along the afferent nerves, it must be in the way in which they end in the central nervous system.

This does not simplify the problem of quality differences. It merely specifies and classifies it. It shows for instance that the nature of the yellow sensation and the white sensation, which we discussed in the last section, is not generically different from that of the red or of the green, and therefore requires no special prominence. These sensations are all problems in brain physiology, and as such have hardly been investigated. Fortunately, since we are concerned with the receptor process alone, we must leave them to one side as beyond our boundaries, but certainly not beyond our interests.

If, as the result of closer study, it has been necessary to ascribe specific quality perception to the central nervous system, the present treatment of the photoreceptor process has shown that at least one property loosely ascribed to the central nervous system can be taken care of by the peripheral receptor process. I refer to the process of adaptation. Among the many relations which an organism maintains with its environment, none would seem so purposeful as its adaptation to a sustained change in the environment. This kind of adaptation belongs to those activities of an animal which have been called its "behavior," and as such, have been considered as remote from analysis in terms of matter and energy as it is possible to get.

Indeed, because of this appearance of purpose, sensory adaptation has been used in the study of animal behavior as a tool with which to prove the presence of a "higher behavior" in animals. It would be beside the point to make specific references to literature of this nature, because of the extent to which it would lead. The photoreceptor process constitutes one of the very few closely studied instances of the behavior of an organized tissue. In the case of *Mya, Ciona,* and *Pholas* it has been necessary to include in this study the sensory behavior of the animal as a whole. And as this chapter has shown, such a use of adaptation is wholly unwarranted. It must by now be clear that adaptation to sustained illumination is essentially a sensory phenomenon, and takes place as the result of a rearrangement in the concentrations of a group of chemical substances and processes which constitute the essential parts of the sense-organ.

Adaptation possesses a property of prime significance, in that it is automatic. Not only does adaptation take place in the sense-organ, away from nervous activity, but it is actually determined by the outside source of stimulation. The equation which describes the situation in the sense-cell during adaptation is the stationary state equation

$$KI = \frac{x^2}{a - x} \qquad [17]$$

which we derived on page 248. In this equation the only two variables are the intensity I of the light and the concentration x of the decomposition products P and A produced by it from the sensitive substance S. Of these two variables it is the intensity I which changes independently, and with it change the concentrations of S, P, and A. Given a value of I, and the equation is at once determined. In short, adaptation is a phenomenon which is controlled not by the animal but by the light. We have heretofore applied the term adaptation in the wrong place as far as sensory adaptation is concerned. We have assumed that the animal adapts itself to the light, whereas, once given the sensory mechanism, we should say that the light adapts the animal to itself.

It is well to look a little more closely at equation [17], and equations [15] and [16], from which it is derived. It is apparent that a change in intensity I will produce an effect on the concentrations of the materials in the sensory system. If the intensity is maintained, however, the system comes into equilibrium. Thus, between the nervous system of the animal and the outside environment there is interposed a sensory system which acts so as to transmit any rapid changes in the environment, but which hides any long-maintained conditions.

This becomes even more evident at different levels of adaptation. In order to produce a given effect, the amount of decomposed materials P and A formed is constant, no matter what the adapting light happens to be. Note carefully that the actual amount of outside light necessary to produce this constant photochemical effect varies strikingly with the intensity, as the values of $\Delta I/I$ in intensity discrimination show (cf. Tables 7 and 8 on pages 249 and 250). We speak of this as a variation in sensitivity. The impulse from the sense-organ, however, is the same and results in a given intensity discrimination. The sense-organ in this way acts as a buffer, in that it changes these variable amounts of environmental disturbance into a constant photochemical disturbance which constitutes the impulse for a given sensory effect. Looked at from this point of view, sense-organs possess a function which is usually not recognized as theirs. The sensory system of animals is not only the traditional receptor system with which we have so far been concerned but is, as well, a protecting layer which stabilizes and buffers the relation between the nervous system and the environment.

REFERENCES

1. Abney, W. deW. Researches in color vision. London, 1913.
2. Adrian, E. D. The all-or-none principle in nerve. *J. Physiol.*, 1913-14, **47**, 460.
3. Adrian, E. D., & Bronk, D. W. The discharge of impulses in motor nerve fibres. Part I. Impulses in single fibres of the phrenic nerve. *J. Physiol.*, 1928, **66**, 81.
4. Adrian, E. D., & Matthews, R. The action of light on the eye. Part II. The process involved in retinal excitation. *J. Physiol.*, 1927, **64**, 279.
5. Adrian, E. D., & Zotterman, Y. The impulses produced by sensory nerve endings. Part III. Impulses set up by touch and pressure. *J. Physiol.*, 1926, **61**, 465.
6. Aubert, A. Physiologie der Netzhaut. Breslau, 1865.
7. Best, F. Ueber die Dunkeladaptation der Netzhaut. *Arch. f. Ophth.*, 1910, **76**, 146.
8. Blaauw, A. H. Die Perception des Lichtes. *Rec. trav. bot. Néerl.* 1909, **5**, 209.
9. Blanchard, J. The brightness sensibility of the retina. *Phys. Rev.*, 1918, Series 2, **11**, 81.
10. Bloch, A. M. Expériences sur la vision. *C. r. Soc. biol.*, 1885, Series 8, **2**, 493.

11. Blondel, A., & Rey, J. Sur la perception des lumières brèves a la limite de leur portée. *J. de physique,* 1911, Series 5, **1**, 530.

12. Boll, F. Zur Anatomie und Physiologie der Retina. *Monatsber. Berlin. Akad.,* 1876, 783.

13. Braunstein, E. P. Zur Lehre von den kurzdauernden Lichtreizen der Netzhaut. *Zsch. f. Sinnesphysiol.,* 1923, **55**, 185.

14. Brown, T. G. On the question of fractional activity ("all or none" phenomenon) in mammalian reflex phenomena. *Proc. Roy. Soc. London* (*B*), 1913-14, **87**, 132.

15. Bunsen, R., & Roscoe, H. Photochemische Untersuchungen IV. *Ann. d. Physik,* 1862, **117**, 529.

16. Crozier, W. J. Wave length of light and photic inhibition of stereotropism in *Tenebrio* larvae. *J. Gen. Physiol.,* 1923-24, **6**, 647.

17. Dobrowolsky, W. Ueber binoculare Farbenmischung. *Pflüg. Arch. f. d. ges. Physiol.,* 1875, **10**, 56.

18. Exner, F. Zur Kenntniss der Grundempfindungen im Helmholtzschen Farbensystem. *Sitzber. Akad. Wiss., Wien, Math.-Naturwiss. Kl., Abt.* II*a*, 1920, **129**, 27.

19. Fechner, G. T. Ueber ein wichtiges psychophysiches Grundgesetz und dessen Beziehung zur Schätzung der Sterngrössen. *Abh. k. sachs. Ges. Wiss., Math.-Phys. Kl.,* 1858, **4**, 31, as published separately; 1859, **4**, 455, as published in bound volume.

20. Forbes, A., & Gregg, A. Electrical studies in mammalian reflexes. II. The correlation between strength of stimuli and the direct and reflex nerve response. *Amer. J. Physiol.,* 1915-16, **39**, 172.

21. Garten, S. Ueber die Veränderungen des Sehpurpurs durch Licht. *Arch. f. Ophth.,* 1906, **63**, 1.

22. Goldberg, E. Beitrag zur Kinetik photochemischer Reaktionen. Die Oxydation von Chinin durch Chromsäure. *Zsch. f. phys. Chem.,* 1920, **41**, 1.

23. Hecht, S. The photochemical nature of the photosensory process. *J. Gen. Physiol.,* 1919-20, **2**, 229.

24. ————. The dark adaptation of the human eye. *J. Gen. Physiol.,* 1919-20, **2**, 499.

25. ————. The nature of foveal dark adaptation. *J. Gen. Physiol.,* 1921-22, **4**, 113.

26. ————. Intensity discrimination and the stationary state. *J. Gen. Physiol.,* 1924, **6**, 355.

27. ————. The visual discrimination of intensity and the Weber-Fechner law. *J. Gen. Physiol.,* 1924-25, **7**, 235.

28. ————. The effect of exposure period and temperature on the photosensory process in *Ciona*. *J. Gen. Physiol.,* 1925-26, **8**, 291.

29. ————. The relation of time, intensity and wave-length in the photosensory system of *Pholas*. *J. Gen. Physiol.,* 1927-28, **11**, 657.

30. ————. Photochemistry of visual purple. I. The kinetics of the decomposition of visual purple by light. *J. Gen. Physiol.,* 1920-21, **3**, 1.

31. ————. II. The effect of temperature on the bleaching of visual purple by light. *J. Gen. Physiol.,* 1920-21, **3**, 285.

32. ————. III. The relation between the intensity of light and the rate of bleaching of visual purple. *J. Gen. Physiol.,* 1924, **6**, 731.

33. ————. The kinetics of dark adaptation. *J. Gen. Physiol.,* 1926-27, **10**, 781.

34. ————. The relation between visual acuity and illumination. *J. Gen Physiol.,* 1928, **11**, 255.

35. ————. On the binocular fusion of colors and its relation to theories of color vision. *Proc. Nat. Acad. Sci.*, 1928, **14**, 237.

36. Hecht, S., & Williams, R. E. The visibility of monochromatic radiation and the absorption spectrum of visual purple. *J. Gen. Physiol.*, 1922, **5**, 1.

37. Hecht, S., & Wolf, E. The visual acuity of the honeybee. *J. Gen. Physiol.*, 1929, **12**, 727.

38. Helmholtz, H. Handbuch der physiologischen Optik. (1st. ed.) Hamburg, Leipzig, 1866.

39. ————. Treatise on physiological optics. (Ed. by J. P. C. Southall.) Ithaca, N. Y., 1924.

40. Hering, E. Der Raumsinn und die Bewegungen in des Auges. Hermann's Handbuch der Physiologie, III, 343. Leipzig, 1879.

41. ————. Grundzüge der Lehre vom Lichtsinn, Berlin, 1920.

42. Hess, C. Neue Untersuchungen über den Lichtsinn bei wirbellosen Tieren. *Pflüg. Arch. f. d. ges. Physiol.*, 1910, **136**, 282.

43. Ives, H. E. The transformation of color mixture equations from one system to another. II. Graphical aids. *J. Franklin Instit.*, 1923, **23**.

44. Judd, D. B. The reduction of data on mixture of color stimuli. *Bull. Amer. Phys. Soc.*, 1929, **4**, 5.

45. Koenig, A. Ueber den menschlichen Sehpurpur und seine Bedeuting für das Sehen. *Sitzber. Akad. Wiss., Berlin,* 1894, 577.

46. ————. Ueber die Anzahl der unterscheidbaren Spectralfarben und Helligkeitsstufen. *Zsch. f. Psychol. u. Physiol. d. Sinnesorg.*, 1895, **8**, 375.

47. ————. Die Abhängigkeit der Sehschärfe von der Beleuchtungsintensität. *Sitzber. Akad. Wiss., Berlin,* 1897, 559.

48. Koenig, A., & Brodhun, E. Experimentelle Untersuchungen über die psychophysische Fundamentalformel in Bezug auf den Gesichtsinn. *Sitzber. Akad. Wiss., Berlin,* 1889, 641.

49. Koenig, A., & Dieterici, C. Die Grundempfindungen in normalen und anomalen Farbensystemen und ihre Intensitätsverteilung im Spektrum. *Zsch. f. Psychol. u. Physiol. d. Sinnesorg.*, 1892, **4**, 241.

50. Koettgen, E. & Abelsdorff, G. Absorption und Zersetzung des Sehpurpurs bei den Wirbeltieren. *Zsch. f. Psychol. u. Physiol. d. Sinnesorg.*, 1896, **12**, 161.

51. Kohlrausch, A. Untersuchungen mit farbigen Schwellenprüflichtern über den Dunkeladaptationsverlauf des normalen Auges. *Pflüg. Arch. f. d. ges. Physiol.*, 1922, **196**, 113.

52. Kries, J. von Die Gesichtsempfindungen. Nagel's Handbuch der Physiologie des Menschen, III, 109. Braunschweig, 1905.

53. Kuehne, W. Chemische Vorgänge in der Netzhaut. Hermann's Handbuch der Physiologie, III, Pt. 1, 235. Leipzig, 1879.

54. Kundt, A. Ueber den Einfluss des Lösungsmittels auf die Absorptionsspectra gelöster absorbirender Medien. *Ann. Physik u. Chem.*, 1878, **4**, 34.

55. Ladd-Franklin, C. The nature of the color sensations. In Helmholtz' Treatise on Physiological Optics. Ithaca, N. Y., 1924.

55*a*. Langley, J. N. On the union of cranial autonomic (visceral) fibers with the nerve cells of the superior cervical ganglion. *J. Physiol.*, 1898, **23**, 240.

56. Lasareff, P. Ionentheorie der Reizung. Bern, Leipzig, 1923.

57. ————. Untersuchungen über die Ionentheorie der Reizung. VI. Ueber die Empfindung der Lichtintensität beim peripheren Sehen auf Grund der Ionentheorie. *Pflüg. Arch. f. d. ges. Physiol.*, 1923, **199**, 290.

58. ————. Zur Theorie der Adaptation der Netzhaut beim Dämmerungssehen. *Pflüg. Arch. f. d. ges. Physiol.*, 1914, **155**, 310.

59. ————. Ueber die Theorie der Dunkeladaptation bei der starken Vorbelichtung. *Pflüg. Arch. f. d. ges. Physiol.*, 1926, **213**, 256.

60. Laurens, H., & Hooker, H. D. Jr. Studies on the relative physiological value of spectral lights. II. The sensibility of *Volvox* to wave-lengths of equal energy content. *J. Exper. Zoöl.*, 1920, **30**, 345.

61. Loeb, J. Forced movements, tropisms, and animal conduct. Philadelphia, 1918.

62. Loeb, J., & Wasteneys, H. The relative efficiency of various parts of the spectrum for the heliotropic reactions of animals and plants. *J. Exper. Zoöl.*, 1916, **20**, 217.

63. Lucas, K. The "all-or-none" contraction of the amphibian skeletal muscle fibre. *J. Physiol.*, 1908-9, **38**, 113.

64. Mast, S. O. The relation between spectral color and stimulation in the lower organisms. *J. Exper. Zoöl.*, 1917, **22**, 471.

65. Müller, J. Handbuch der Physiologie des Menschen. Coblenz, 1840.

66. Nagel, W. Adaptation, Dämmerungssehen und Duplizitätstheorie. Helmholtz' Handbuch der physiologischen Optik, II, 264. Hamburg, 1911.

67. Obreshkove, V. The photic reactions of tadpoles in relation to the Bunsen-Roscoe law. *J. Exper. Zoöl.*, 1921, **34**, 235.

68. Piéron, H. Les lois de l'excitabilité lumineuse des lamellibranches. Récherches sur la loi de variation des énergies liminaires chez *Mya arenaria*. *Bull. Stat. Biol. Arcachon,* 1926, **23**, 1.

69. ————. De la variation de l'énergie liminaire en function de la durée d'excitation pour la vision fovéale. *C. r. Acad. Sci.*, 1920, **170**, 525.

70. Piper, H. Ueber Dunkeladaptation. *Zsch. f. Psychol. u. Physiol. d. Sinnesorg.*, 1903, **31**, 161.

71. Pütter, A. Studien zur Theorie der Reizvorgänge. Parts I-IV. Mitteilung. *Pflüg. Arch. f. d. ges. Physiol.*, 1918, **171**, 201.

72. ————. VII. Das Abklingen der Erregung. *Pflüg. Arch. f. d. ges. Physiol.*, 1920, **180**, 260.

73. Reeves, P. Rate of pupillary dilatation and contraction. *Psychol. Rev.*, 1918, **25**, 330.

74. ————. The effect of size of stimulus and exposure time on retinal threshold. *Astrophys. J.*, 1918, **47**, 141.

75. Rochat, G. F. Étude quantitative du fusionnement binoculaire des couleurs complementaires. *Arch. Néerl. d. physiol.*, 1922, **7**, 263.

76. Rutenberg, D. Ueber die Netzhautreizung durch kurzdauerende Lichtblitze und Lichtlücken. *Zsch. f. Sinnesphysiol.*, 1914, **48**, 268.

77. Trendelenburg, W. Quantitative Untersuchungen über die Bleichung des Sehpurpurs in monochromatischem Licht. *Zsch. f. Psychol. u. Physiol. d. Sinnesorg*, 1904, **37**, 1.

78. ————. Weitere Versuche über binokulare Mischung von Spektralfarben. *Pflüg. Arch. f. d. ges. Physiol.*, 1923, **201**, 235.

79. Troland, L. T. Report of the Colorimetry Committee of the Optical Society of America, 1920-21. *J. Opt. Soc. Amer.*, 1922, **6**, 527.

80. Uhthoff, W. Ueber das Abhängigkeitsverhältniss der Sehschärfe von der Beleuchtungsintensität. *Arch. f. Ophthal.*, 1886, **32**, Abt. 1, 171.

81. Weber, E. H. De pulsu, resorptione, auditu et tactu annotationes anatomicae et physiologicae. Author's summary: Ueber den Tastsinn. *Arch. Anat. u. Physiol.*, 1835, 152.

82. Weigert, F. Ein photochemisches Modell der Retina. *Pflüg. Arch. f. d. ges. Physiol.*, 1921, **190**, 177.

83. Young, T. On the theory of light and colors. Lectures in natural philosophy. II, 613. London, 1807.

CHAPTER 6

HEARING I

H. Banister

Cambridge University

I. Sounds

When we say we hear, we mean that we experience sensations which have been produced by the excitation of the auditory mechanism, excluding abnormalities such as hallucinations. These elementary experiences, by means of which we are aware of so much with regard to the external world, by which we can receive the thoughts of others, and which contribute so largely to our aesthetic enjoyment of life through poetry and music, are called sounds. Yet such sounds have no objective existence; outside ourselves all is quietness. Externally there are simply bodies in vibration. If these vibrations are communicated to the ear, they may, provided they lie within certain limits of amplitude and frequency, be perceived as sounds. Unfortunately for clarity in thinking, the word "sound" is used in an objective way as though the sound existed as a sound independently of ourselves. We say, "Did you hear that sound?" when a gun is fired, while we should say, "Did you hear that explosion?" for it is the explosion caused by the firing of the cartridge which is perceived; externally to ourselves are waves of compression and rarefaction in the air, nothing more. But custom is strong, so with this warning the word "sound" will be used in what follows in the two ways in which it is commonly used (*a*) for the subjective experience, and (*b*) for the origin of the source of that experience.

Physics of Sound. Sound waves are initiated by bodies in vibration, and for their transmission some material substance is necessary. An electric bell in a hermetically sealed glass jar can be seen and heard to ring; if the air is pumped out it can still be seen to ring, but is no longer heard.

The vibrations are longitudinal, that is, in the direction in which the waves are travelling; they are strictly analogous to the closing in and opening out of the carriages of a train when shunting operations are in progress. The adjacent particles of the air, or other substance, through which the sound waves are travelling, are alternately compressed and expanded; the wave advances, but the individual particles remain relatively stationary, moving to and fro about a mean position. The same kind of phenomenon is to be observed with water waves. A boat at sea rises and falls as a wave passes it; the wave travels on, but the boat and the water remain. The boat and the water, however, move up and down in a direction *at right angles* to that in which the wave is travelling; with sound waves the movement is backwards and forwards *in* the direction in which the wave travels.

Though the direction of the movement of the particles of the medium which transmits sound waves is along the direction of travel, it is customary to represent the sound wave as though it were a water wave (see Figure 1). This may cause a little difficulty when one endeavors to visualize what is happening. It is to be remembered that where the wave crest is shown, e.g., at *A* and *E*, the particles are closest together; at the trough, *B* and *F*, they are widest apart; and at *C* and *D* they are in their normal state of proximity to their neighbors. Waves may vary in amplitude of vibration (*Aa*) and in wave-length (*AE*). The greater the amplitude, the more powerful the waves. Since waves travel with the same velocity in a uniform medium,[1] the frequencies with which waves of different wave-lengths pass a given point will vary. Waves of long wave-lengths will be less frequent than waves of short wave-length, and there will be strict proportionality between the wave-length and the frequency. In symbols, $\frac{\lambda}{f}$ is constant for all waves in the same medium, where λ is the wave-length and f is the frequency with which the waves succeed one another.

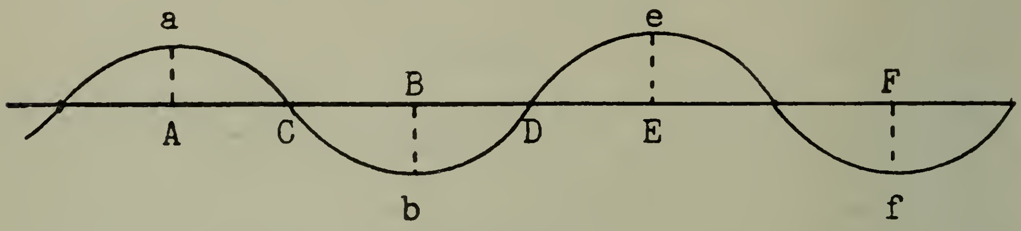


FIGURE 1

The velocity of travel of sound waves varies with the medium. It is about 330 meters per second in air, as compared with about 520 meters in glass and 1435 meters per second in water.

The vibrations may be of innumerable kinds: they may be simple to-and-fro movements similar to the swing of a pendulum, as represented in Figure 1 (which tracing might have been obtained by moving the paper along in a direction at right angles to the direction of the swing of the pendulum bob, to which a marker had been attached); or they may be more complicated as shown in Figure 2; or more complicated still as in Figure 3.

The simplest kind of vibration, that of the bob of a long pendulum when the amplitude of the swing is small, is called "simple harmonic motion." It will be noticed that the curve of Figure 2 is made up of two simple harmonic curves. This is an example of a general law, first formulated by Fourier, that any periodic motion, however complicated, can

---

[1] There are exceptions, for example, when waves are conveyed through air in tubes and when they are due to explosions.

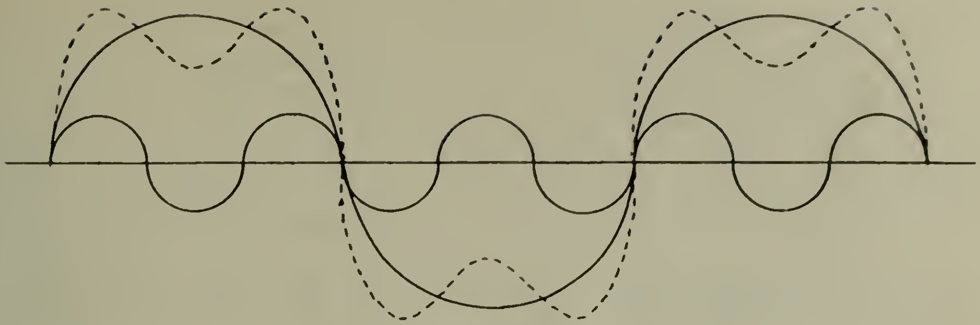


FIGURE 2
A Compound Wave Made Up of Two Simple Harmonics

be analyzed into components, each one of which is simple harmonic. This is a very important law for students of hearing. It enables us to explain, on physical grounds, variations in "timbre."

Sound waves suffer reflection, as evidenced in echoes; and if the medium through which they travel is not uniform, the waves are refracted (bent): cold air, for example, transmits sound waves more slowly than hot, with the result that sounds are heard at much greater distances on cold than on hot days—on cold days the waves are refracted down towards the earth where the atmosphere is cold and travel along the surface; on hot days they are bent upwards and pass over our heads.

Besides reflection and refraction, sound waves, owing to their long wave-length, can bend round corners even though they travel in a uniform medium—they suffer what is called diffraction. It is owing solely to its very short wave-length that light travels in straight lines; with appropriate methods light waves can be shown to bend round corners. In water waves, diffraction is easily observed; a boulder is an impassable obstacle to small ripples; but, with a wave of long wave-length, the water moves up and down equally on both sides of the boulder —at the far side the water rises and falls as it would if the boulder were not there. Similarly with sound waves. They bend round corners. With waves of long wave-length and small frequency, the wave motion is just as great at the far ear as at the near; but as the wave-length becomes shorter the far ear is shielded more and more by the head, till,

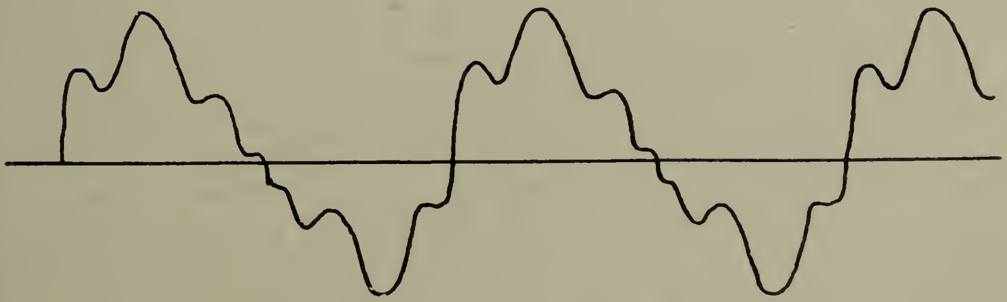


FIGURE 3
A More Complicated Wave Which Can Be Analyzed into Simple Harmonic Waves

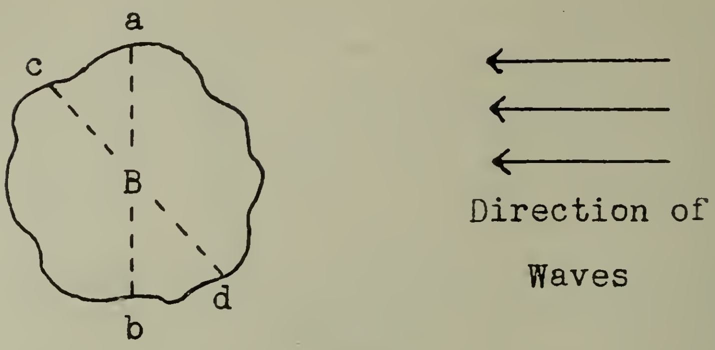


FIGURE 4

with very high frequencies, there is possibly no wave motion at all at the far side.

If we study the wave motion round a boulder (*B*) in the sea, we notice that the waves rise and fall synchronously at opposite points (*a* and *b*), where the line joining them is at right angles to the direction in which the waves are travelling; while at other opposite points, e.g., *c* and *d*, there is, as a rule, no correspondence; a crest at *d* will correspond with a trough or any other state at *c*, depending on the wave-length and the size of the boulder. This difference of level, which obviously depends on the amplitude of the wave and also on the wave-length, is spoken of as a difference of "phase,"[2] the difference of phase taking into account the wave-length and the amplitude. For purposes of understanding, however, a phase difference may be regarded as a difference of level.

II. TONES

Tone. Suppose that we have a body which can vibrate with all possible frequencies from 0 to 40,000 or 50,000 complete vibrations (from any given position, to the extreme in one direction, back to the extreme in the other, and then on to its initial position) per second; that the body vibrates with simple harmonic motion; and that the amplitude of the vibration is adequate (we shall discuss this adequacy below). Then when the frequency is very low, nothing will be perceived; as the frequency is increased, a "flutter" will be observed, which is localized in the tympanic membrane. This "flutter" seems to be due to the vibration of the tympanic membrane and perhaps to a lesser extent, to the arousal of tactile sensations in the walls of the external auditory meatus. It is clearly a tactual, not an auditory sensation, and is probably due to the excitation of the nerve-fibers, with which the tympanic membrane and the external auditory meatus are furnished. On further increasing the frequency, an auditory sensation is aroused, which may be accompanied by the tactual sensation. This auditory sensation is at first very faint and has no discernible pitch; it is a noise. Gradually, however, as the frequency is increased, the tactile sensation may disappear, and the sound takes on the quality of a tone of very low pitch. The tone now passes through the whole gamut of tones, eventually becoming a very pronounced "hiss," again with no determinable pitch, before fading rather suddenly away as the frequency becomes too high for auditory perception. Except at the limits, where we hear noises, these sounds are pure tones.

---

[2] This holds strictly only when the motion is "simple harmonic."

Tone Character. A perfectly pure tone probably never exists in nature; but, with reasonably pure tones, there are marked differences in quality—low tones appear voluminous and massive, high tones thin and piercing; low tones again have a *u* vowel quality; high tones an *i* quality.

Limits of Hearing. Much work has been done with the aim of determining the limits of the frequencies which produce auditory sensations. The lower limit has been variously estimated from 9 to 12 d.v./sec. (Appunn, in 1887) to 26 d.v./sec. (1). The determination is very difficult, for not only is it dependent on the discrimination between an auditory and a tactile sensation but also, owing to the difficulty of obtaining absolutely pure tones, it is always possible that the auditory stimulus perceived was not the one it was thought to be (an overtone might, for example, be mistaken for the fundamental) especially since frequencies less than 40 d.v./sec. lose their tone character. Another factor affecting the results, and one which was not fully recognized at first, is that at the auditory limits the magnitude of the vibration must be nicely adjusted, as we shall see later (20, 28), if an auditory sensation is to be produced. Recent research places the lower limit somewhere in the neighborhood of 20 d.v./sec. with large individual variations. The upper limit of hearing, which was placed at about 200,000 d.v./sec. by the earliest investigators, is now generally considered to be not more than 20,000 to 25,000 d.v./sec. The upper limit shows a considerable lowering with age. In music, the lowest note of the organ is about 16.5 d.v./sec., and the highest note of the piccolo, 4752 d.v./sec.

Pitch. As we have seen, the pitch of a pure tone varies with the frequency of the vibrations, if we exclude the frequencies at the extremes which have no pitch. Tones of low pitch have few, tones of high pitch have many vibrations every second.[3] It is customary, therefore, to designate the pitch of a tone by the frequency. But this is not an exact description, for many people hear the same tone as being of different pitch with the two ears (diplacusis) (24, 42). This may occur throughout the whole tonal range, though it is often confined to certain frequencies. The difference in pitch between the two ears may amount to a semitone, and in experiments on localization the pitch may form the basis of the judgment of the apparent source of sound. Also the pitch of a low tone may vary with its loudness, becoming lower as the loudness increases (26).

---

[3] It is difficult to assign any reason for the use of the words "high" and "low" in connection with pitch. Perhaps it is because the chest is the resonator used principally when singing low notes, while the resonators for high notes are in the head. This may be one reason why an indifferent singer tends to stand on tip-toe when endeavoring to reach a note which is a little beyond his compass.

Pitch Discrimination. The smallest perceptible difference of pitch, the differential threshold, varies with the frequency of the tone. At the upper and lower limits, discrimination is very poor, but in the middle ranges discrimination is good and can be greatly improved with practice. Stumpf (50), for example, could discriminate differences of 0.25 of a vibration at 200 d.v., of 0.28 at 400 and of 0.24 at 600 d.v. This discrimination is very fine. It is remarkable too in that the differential threshold is not in agreement with Weber's law, i.e., the difference in frequency is not proportional to the actual frequency. On this point the findings agree with those published by Luft (29) in 1888. He obtained thresholds of 0.23 at 256 d.v., 0.25 at 512 d.v., and 0.22 at 1024 d.v. These and other findings have given rise to the statement that pitch discrimination is perhaps the one sensory function which does not obey Weber's law. But later investigations (49, 54, 22) indicate that probably Weber's law holds for differences of frequency in the middle range. In the figure, the data of the two latter experimenters are shown. Vance (54) experimented with tuning forks; Knudsen (22), with tones produced by an electrical oscillating circuit.

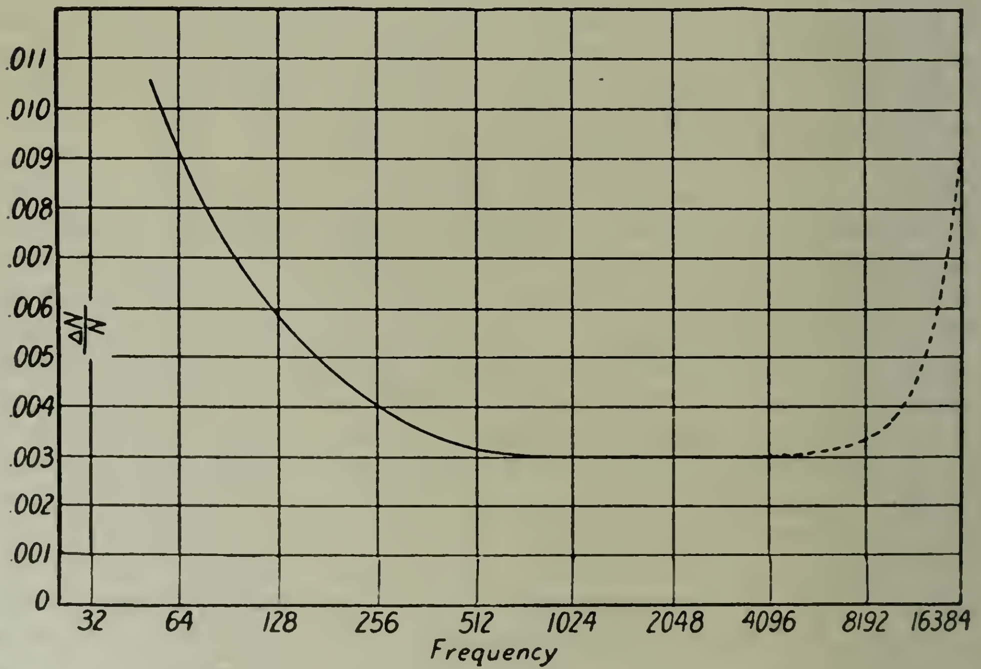


FIGURE 5

(From: Fletcher, H. Physical measurements in audition. *J. Franklin Instit.*, 1923, **196**, No. 3.)

At 256 d.v./sec. the minimal perceptible difference, according to Knudsen, is given by $\Delta N/N = 0.004$.

$$\therefore \Delta N = 0.004 \times 256$$
$$= 1.024$$

The difference limen depends to a certain extent on the intensity of the tone. For sounds of low intensity the limen is considerably higher than for normal intensities. The maximum sensitivity is obtained when the physical intensity is not less than about 10^6 times the intensity of the absolute limen.

It is easier to dicriminate differences of pitch with binaural than with monaural hearing, and also easier to tell that two notes differ in pitch than to say which is the higher.

Minimum Number of Vibrations for the Perception of Tone. The minimum number of vibrations necessary to produce the sensation of tone depends on the amplitude of the vibrations and also on the pitch of the tone. With tones of moderate intensity in the middle range, two vibrations are said to suffice, though more are necessary for the perception of the pitch of the tone. But such determinations are extremely hard to make and must be accepted with reserve. Although only two vibrations are initiated, as in Kohlrausch's experiments where two teeth attached to a pendulum struck a card, it is impossible to insure that the ear does not receive more than two waves, for there may be reflection of the wave to the ear from neighboring objects.

Bode (9) gives the following as being the minimum number of vibrations necessary for the determination of pitch. Too few vibrations produce a sensation of noise.

Freq. d.v.	Number of vibrations necessary for pitch determination	
	Weak intensity	Medium intensity
128	12.1	—
256	—	17.6
384	24.08	17.1
512	29.64	21.8

If the tones are of very low intensity, many more vibrations are necessary; two or more seconds even may be required for the perception of their pitch.

Tonal Islands and Gaps. Gaps have often been found in the range of frequencies which can be heard. Stewart (48) reports that he finds such gaps in most subjects. Bunch (10) also considers that they are more common than is ordinarily thought. But others suggest that these "gaps" may be only regions of reduced sensibility; that the tones can be heard if their intensity is sufficiently increased. At the same time there is another alternative. Tones when increased in loudness become more noisy, and it may be that it is the noise which is heard and that the particular tones themselves are not heard.

Loudness. What is the effect when we have a body vibrating in the simplest possible manner—executing simple harmonic motion with a frequency in the middle of the audible range—and when the amplitude of the vibrations is gradually increased from zero? When the amplitude of the vibrations is extremely small, we are not aware of the fact that the body is vibrating; but as the amplitude is increased there comes a stage when we perceive that something is being heard, but cannot tell what. The sound has no pitch. It is a very faint indeterminate sound, more accurately described perhaps as a noise, which gradually acquires tone character. The tone then increases in loudness, becoming at the same time more noisy. At a certain intensity, a sensation appears which becomes acutely painful as the loudness is still further increased.

It would seem that the latter sensation is extra-auditory. It is observable at approximately the same intensity by abnormal and normal ears. This intensity is about that required to excite the tactile nerves in the finger tips (19). It is presumed, therefore, that the sensation is tactual, being due to the activation of the nerves of the tympanic membrane, etc.

The intensities necessary to produce the auditory and tactile sensations vary with the frequency. In Figure 6 are shown the variations in pressure which, it is suggested, produce these sensations. The abscissa gives the frequency, and the ordinate, the estimated root mean square change in pressure on the ear-drum. For calibration purposes, the pressure due to the air vibrations in a chamber of approximately the same volume as the external auditory meatus was measured. Since these pressures were obtained on the assumption that the external auditory meatus and the tympanic membrane are rigid, they are approximations only.

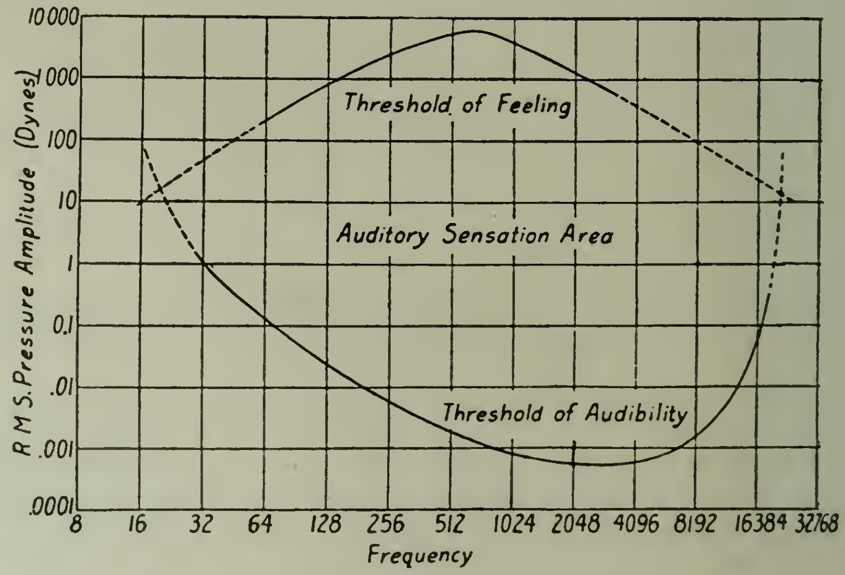


FIGURE 6

(From: Fletcher, H. Physical measurements in audition. *J. Franklin Instit.*, 1923, **196**, No. 3.)

Minimum Intensity for Audition. Extreme difficulty is found in determining the absolute threshold for audition, i.e., the least physical stimulus which will produce an auditory sensation. Since the physical stimulus is, in general, due to vibrations in the air, it may be measured in various ways. The earliest experimenters expressed their results in terms of the amplitude of the vibration. Toepler and Boltzman (1870) found that the minimum amplitude was 200×10^{-8} cm. for a tone of 180 d.v./sec. from an organ pipe; Rayleigh found that an amplitude of 12.7×10^{-8} cm. was sufficient for a 256 d.v. tuning fork, while, for a frequency of 2730 d.v./sec. obtained from a whistle, this amplitude was reduced to 8×10^{-8} cm. The present tendency is to express the intensity in terms of the root mean square of the change in pressure on the tympanic membrane (dynes per sq. cm.), as has already been indicated. The average sensitivity for normal ears varies from .15 dynes/cm.^{2} at 60 d.v. to .001 dynes/cm.^{2} at 1000 d.v., at which value it remains constant up to about 4000 d.v. Thereafter the sensitivity rapidly decreases. All such results must be considered as approximations only. In this field, more than in most, are individual differences and variations in the same individual during the course of an experiment to be found, while the effects of different methods of presentation of the stimulus are most marked. It appears that the automatization of the responses produces a lower threshold and more consistent results; while distraction of attention by other mental occupation, such as reading or working out problems, also results in a lowered threshold (51). These liminal sensations also take time to develop, and many of the variations found may be due to the difficulty in recognizing the sound. Practiced observers often report that before the point of recognition is reached some effect of a sensory nature is certainly produced, which may, for example, be thermal or tactile in character. There is something, they maintain, which is very different from positive silence, though no auditory stimulus is perceived. Such sensations, as has already been indicated, probably have their origin outside the inner ear. They are most noticeable in a slightly deaf ear (47, 20). In fact there would seem every reason to suppose that most auditory stimuli produce tactile as well as auditory sensations. It is only when the former become excessive that they are noticeable under ordinary circumstances.

The question whether the threshold for two-eared hearing is lower than that for monaural hearing is still unsettled.

Maximum Intensity for Audition. In Figure 6 the upper curve shows the average pressure (48 ears) which was required to produce a tactile sensation in addition to the auditory sensation. In the middle range, the sensation is one of pricking. It becomes acutely painful as the intensity is further increased.[4] The curve of pain is very indefinite. For lower frequencies, the sensation is milder; at about 60 d.v/sec. it becomes a flutter. For lower frequencies still, the auditory and tactile sensations become indistinguishable. The maximum intensity for audition may well be taken as that at which the tactile sensation can just be perceived.

---

[4] "Even a deaf mute, though he has no sensation of tone, is able to feel such excessive stimulation" (23).

It has been suggested that the upper and lower frequency limits for audition be taken at the two points at which the minimum and maximum audibility curves intersect. These, on extrapolation from present data, would give the limits at about 20 d.v./sec. and 24,000 d.v./sec.

Intensity Difference Limen. With modern physical developments, measurement of this limen has become easier, and the data obtained, for example, by Wien, who used a telephone which was placed near the mouth of a Helmholtz resonator, and by Deenik, who worked with tuning forks and organ pipes, have been found to be more accurate than might be expected.

In the investigations of Knudsen (24) and Riesz (43) the currents from two oscillating circuits were led to a telephone through an alternator, which was so arranged that, as the connection between one circuit and the telephone was broken, the connection between the second circuit and the telephone was made. Knudsen made these alternations by hand and by motor. Riesz used a motor and found the optimum results were obtained when the change was made about three times a second.

Since both circuits were in tune, the observer heard a continuous tone with fluctuations in loudness, if the two tones differed in this respect. The intensities, which were measured, were varied and the limen obtained.

It was found that Weber's law held, provided the intensity was not less than 10^6 times the minimal audible intensity.

An average of Knudsen's results for four tones of 400, 1000, 2000, and 4000 d.v./sec. is given in Figure 7. Riesz's results agree fairly well. He used seven tones, from 35 to 10,000 d.v./sec., and found that while Weber's law held for each tone it varied from tone to tone, lying between 0.05 and 0.15 as compared with Knudsen's 0.09 and 0.11 for a more limited range.

Riesz calculates that the ear can distinguish about 370 tones of different intensity between the thresholds of audition and feeling at 1300 d.v./sec.

Knudsen and Jones calculate that the normal ear can appreciate about 65 gradations of loudness for a tone of 32 d.v./sec., and that 270 gradations can be appreciated at 512 d.v./sec. For very high as for low tones, the number of gradations between the minimum and maximum limits of audibility are very few.

Total Number of Pure Tone Sensations. Since we know the range of frequencies and intensities which give rise to auditory sensation, it is possible, if we know also the difference limens for frequency and intensity, to determine the total number of pure tone sensations which can be discriminated, each one from every other, in either pitch or intensity, or in both. Such calculations have been made. It appears that, in all, their number is over 300,000—Knudsen puts the number at 324,000.

III. Compound Tones and Noise

Tone and Noise. So far we have considered theoretically pure tones, which have been produced by simple harmonic vibrations. Yet, though the vibrations are simple harmonic, the auditory sensation is not always that usually designated by the description "pure tone." At the limits of audition when the vibrations are too small or too large, or when their period is too slow or too fast, and also when the total number of vibrations are too few (their amplitude and frequency being well within the middle range), the resulting sensation is one of noise. A pure tone whose pitch is varied too rapidly will also produce a sensation of noise, as may a combination of pure tones in certain circumstances, e.g., beats.

All tones normally produced are more or less noisy. It is possible to rank sounds in order of noisiness, the first sound being the most absolute noise and the last the purest tone. If the vibrations producing these various sounds are examined, it is found that, whereas the pure tone is produced by simple harmonic motion of a single frequency, the vibrations producing the absolute noise have no kind of periodicity. But, as we have

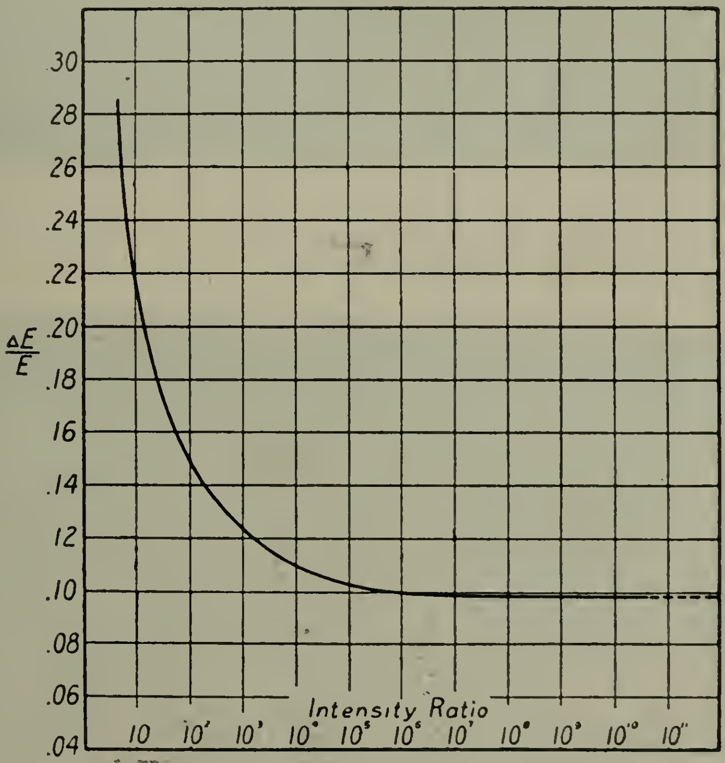


FIGURE 7

Minimum Perceptible Differences in Intensity

(From: Fletcher, H. Physical measurements in audition. *J. Franklin Instit.*, 1923, **196**, No. 3.)

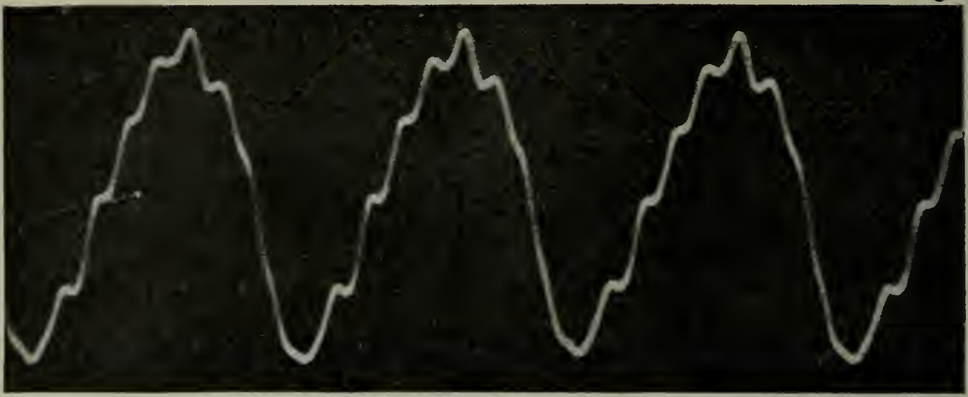


FIGURE 8
(From: Miller, D. C. The science of musical sounds. New York: Macmillan, 1916. P. 196.)

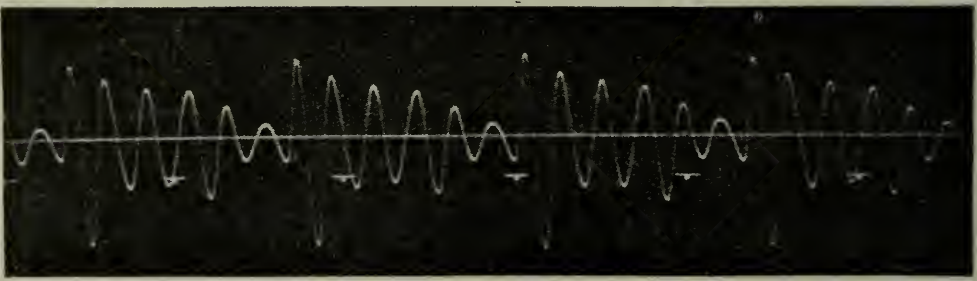


FIGURE 9
(From: Miller, D. C. The science of musical sounds. New York: Macmillan, 1916. P. 203.)

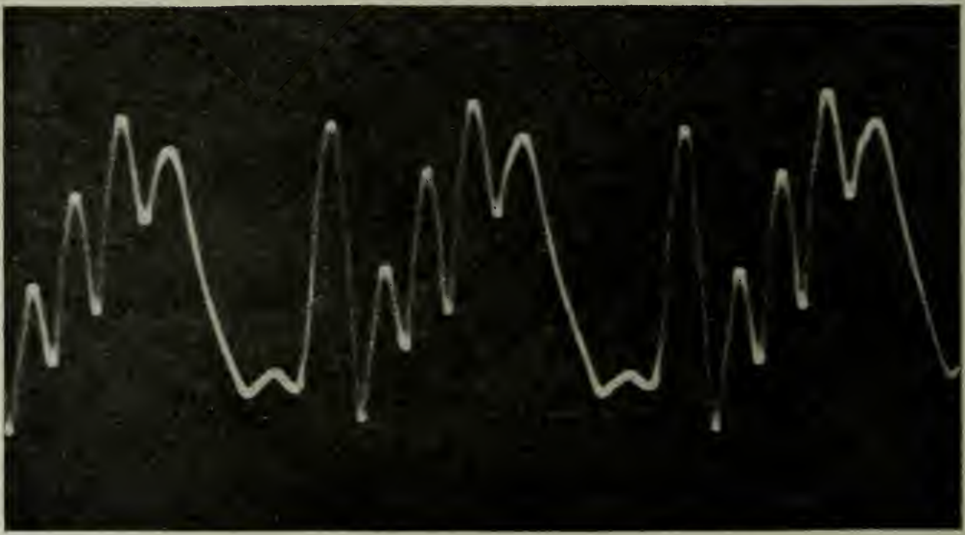


FIGURE 10
(From: Miller, D. C. The science of musical sounds. New York: Macmillan, 1916. P. 200.)

no pure tones, so we have no absolute noises in nature. In all noises, however made, there is always some semblance of pitch. The roar of the surf breaking on the shore, the trickle of the brook, and the rustling of the leaves in the trees, each has a certain tone quality, a certain pitch, however elusive. When the vibrations producing these sounds are examined, they are found to exhibit a certain periodicity, though it may be hidden by much irregularity.

Though there is this gradual transition from tone to noise and it is impossible to decide on any fixed dividing line, yet we each have an absolute impression, however variable, by which we judge sounds and classify them as tones or noises The former are simple, smooth, pleasant; the latter, complex, rough, and unpleasant.[5]

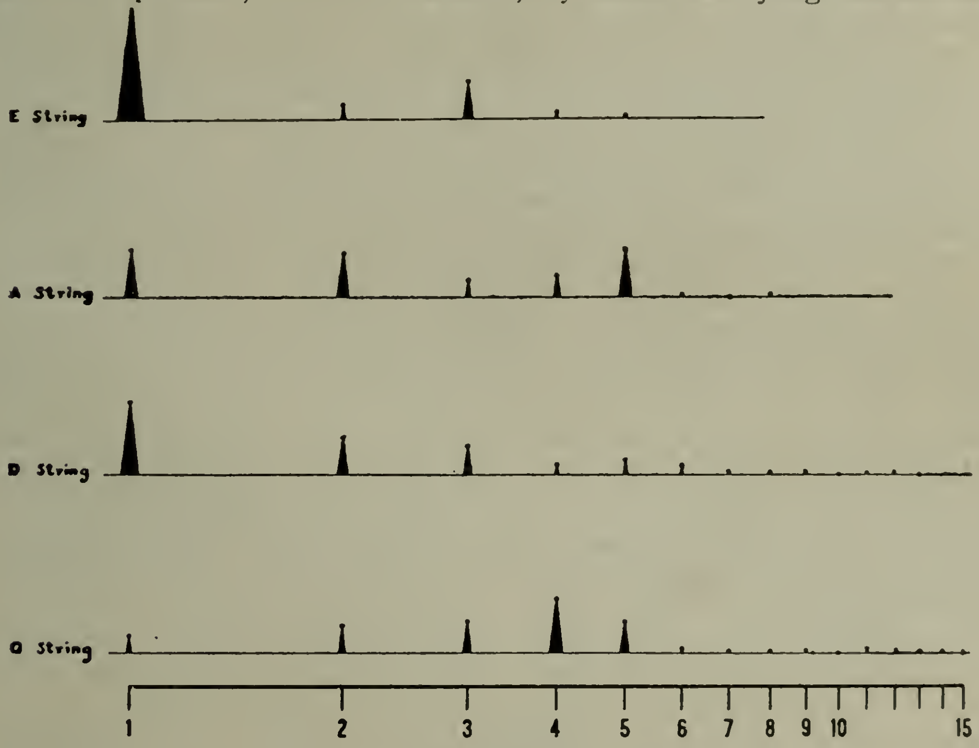


FIGURE 11
Analyses of Violin Tones
(From: Miller, D. C. The science of musical sounds. New York: Macmillan, 1916. P. 198.)

Timbre. A note played by a violin is not mistaken for the same note on the piano or on the trombone; the tones have different qualities according to the instruments by which they are produced. This difference in quality is called "timbre." It depends on the strengths and numbers of the various overtones which are present. Tracings of the vibrations of a violin tone (34, p. 196), of a horn (34, p. 203), and of a clarinet (34, p. 200) are given in Figures 8, 9, and 10. Each of these curves can be analyzed into its constituents, in accordance with Fourier's theorem, and the overtones present and the strength of each can be determined. Such an analysis of the tones produced by the four strings of the violin is shown in Figure 11 (34, p. 198).

---

[5] A child's incomprehensible (to the adult) love of noise is probably due more to his love of *making* the noise than to the pleasure in the auditory sensation *qua* sensation.

The scale at the bottom shows the various overtones up to 15. (For some complex tones as many as 40 overtones may be identified.) It will be seen that the strings have very different characteristics. The fundamental is strongly marked in the E string (this being indicated by the thin triangle over 1), the first overtone is relatively weak, the second is strong, while the rest of the overtones are almost negligible. To be contrasted with this is the vibration of the G string, which has very pronounced overtones, the first, second, third, and fourth being each more powerful than the fundamental. It is owing to these differences between the overtones of the various strings, with their consequent differences in quality, that the violin—and also the 'cello—enjoys such pre-eminence as a solo instrument. A skilled performer by playing on the string he wishes can obtain very great variety of timbre; he can also change the timbre of any given string in innumerable ways by his manner of bowing, etc. It must not be forgotten that the analysis shown is that of a particular violin, and that violins differ greatly among themselves; as a general rule, however, their tones are prominent in the first four overtones.

That the "timbre" of a tone depends on the overtones has been demonstrated in two ways: (*a*) if an electric filter, which eliminates the higher frequencies, is introduced into the receiving circuit of a wireless set, the tone of a violin sounds like that of a tuning fork; and (*b*), conversely, by combining a number of pure tones of the correct frequencies and intensities, a single tone like that of a violin or any other instrument may be produced.

Though a tone may contain many overtones, the latter are seldom perceived by the untrained ear except by their effect on the timbre of the tone. A little training, however, enables the novice to pick out some of the overtones present, for example, in the note given by a monochord. The wire is made to vibrate and, as the tone dies away, the wire is gently touched at the center with a fine paint-brush. This effectively stops the fundamental, and the first overtone is heard. If attention is concentrated on this overtone and the wire made to vibrate again, the overtone will be easily recognized. In the same way, the presence of higher overtones may be demonstrated.

Pitch of Pure and Compound Tone. The frequency of a compound tone, such as we have been discussing, is that of its lowest frequency, but its pitch is usually slightly higher than that of a pure tone of the same frequency. The more pronounced the overtones, especially the higher ones, the higher the pitch appears to be. As a corollary to this, one finds that the pitch of a tuning fork rises as the fork is moved slowly away from the ear; distance is supposed to reduce the strength of the fundamental tone proportionally more than that of the overtones.

Physical Criterion for Determining the Pitch of a Tone. We have already seen that removing the higher frequencies of a complex tone alters the quality but not the pitch (excluding the slight effect noted in the preceding paragraph) of the tone. It has been shown (16) that the lower frequencies also may be eliminated with a similar result. For example, the vowel *ah* sung at a frequency 145 d.v./sec. had the same pitch and fundamental when the fundamental itself and the first, second, sixth, and higher overtones were eliminated; the third, fourth, and fifth overtones alone were sufficient to cause the hearing of the original frequency (145 d.v.). In general, the elimination of the fundamental and of the first two or three overtones had little or no appreciable effect on the quality or pitch of the notes of a rich baritone or contralto voice; but the quality of these voices was noticeably affected if the higher overtones were eliminated—even though all the overtones below the fifteenth were present. With a high soprano, on the other hand, the elimination of the higher overtones affected the quality very little. Thus the elimination of frequencies above 2000 d.v./sec. was found to produce the surprising result of spoiling the quality of the low voice more than that of the high.

The fact that the pitch of a compound tone may be much lower than that of the lowest of its constituents was confirmed synthetically. For example, any three consecutive tones of a series of ten, with frequencies 100, 200, 300, etc., to 1000 d.v./sec., such as 300, 400, 500, or 700, 800, 900, etc., gave a compound tone of the same pitch, viz., 100 d.v./sec., but of varying quality, provided the intensity was not too low. With frequencies of 600, 800, and 1000 the pitch was similarly 200 d.v./sec.

The explanation of these phenomena is to be sought in the non-linearity of the auditory mechanism, owing to which the various difference and summation tones are produced.[6] If the fundamental and first few overtones are eliminated from the air waves which fall on the tympanic membrane, they are reintroduced (though not in their orginal strength) before the waves can excite the hair-cells of the organ of Corti owing to the physical properties of the ear.

Masking. If two forks of different frequencies (the difference of frequency being sufficient to avoid beats[7]) are sounded together, it is found that the higher tone disappears first. If the lower fork is now damped, the higher tone will again be heard. The higher tone is said to have been masked by the lower.

It was at one time thought that a low tone would completely mask higher tones of considerable intensity but could not be masked by the higher tones, however intense. This is not correct. A low tone does not mask a tone of much higher frequency to any marked degree, while a higher tone will mask a lower if the difference in pitch is relatively small and the higher tone is more intense, provided the difference in frequency does not produce beats. Figure 12 from Fletcher (14) shows the masking effect of a tone of 800 d.v./sec. on various other tones.

---

[6] These phenomena are discussed in Section IV.

[7] Beats will be discussed in Section IV.

The loudness[8] of the 800 frequency is shown as abscissa, and the threshold shift, i.e., the number of units of loudness by which any other given tone must be increased for it to be heard in the presence of the masking tone, is represented by the ordinate. For example, if the loudness of the masking tone is 60 units, it will completely mask the other tones as follows:

250	d.v./sec.	if	the	loudness	is	less	than	3	units
600	"	"	"	"	"	"	"	15	"
700	"	"	"	"	"	"	"	32	"
850	"	"	"	"	"	"	"	44	"
1100	"	"	"	"	"	"	"	41	"
1500	"	"	"	"	"	"	"	13	"
2000	"	"	"	"	"	"	"	10	"

According to the data given in the figure, if a complex tone of 800, 600, and 2000 d.v./sec. has loudness values of 50, 10, and 5 units respectively, the 600 frequency will not be heard, since the 800 tone will mask a 600 tone of 13 units of loudness. If the complex tone is amplified so that the loudness of each component is increased by 20 units, the loudness of the components will be 70, 30, and 25 units. Now the 600 frequency will be heard, but the 2000 frequency will have disappeared; since an 800 frequency of 70 units' loudness will not mask a 600 frequency if the loudness is over 18 units, but will mask a tone of 2000 frequency of 30 units of loudness. Hence, we see that amplifica-

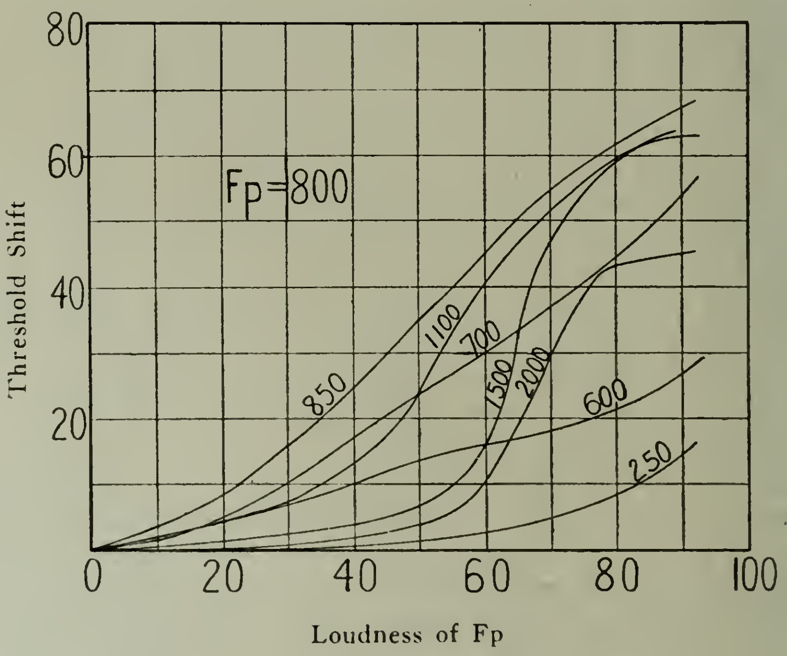


FIGURE 12
(Redrawn from: Fletcher, H. Physical measurements in audition. *J. Franklin Instit.*, 1923, **196**, No. 3.)

---

[8] Loudness is measured in terms of the logarithm of the ratio of the physical intensity to the minimum intensity for audition. An increase of intensity from 10 to 100 units would, on this scale, produce an increase of loudness from one to two units.

tion may change the quality of a complex sound very considerably. This is one factor in the distortion and change in quality caused by loud-speakers.

It is suggested that these facts can be explained on a resonance theory if we allow for the harmonics and combination tones which are introduced through the asymmetrical functioning of the auditory apparatus.

Comparative Loudness of Pure Tones. The differences in quality of tones of high and low pitch make a comparison of their loudness difficult, and masking effects accentuate the difficulty. Helmholtz noticed that of two tones of the same amplitude the higher sounded the louder, and subsequent observers have confirmed his observations. Recently MacKenzie (30) and Kingsbury (23) have endeavored to make direct measurements[9] of the comparative loudness of tones. The former used a flicker method, the tones to be compared being presented to the ears alternately at the rate of about 25 alternations a second. When sounds are presented in this way, the louder tone appears to be interrupted, while the quieter tone seems to be continuous, to sound all the time, but to be drowned when the louder tone is heard. The quieter tone appears to be a continuous background of tone into which the louder tone bursts. As this phenomenal difference appears to vary with intensity only, MacKenzie suggests that the sounds be considered to be equally loud when this distinction cannot be made. This assumption—that the loudnesses of the two tones are equal when both are heard equally; when one tone does not appear to irrupt on to the other which is continuous—must be accepted with care. It is one way of defining loudness but this loudness may or may not be the same loudness as that measured in other ways.

With this definition of loudness it was found that if two tones f_1 and f_2 were judged equally loud, and if another tone f_3, were made equal to that of f_2, then the loudness of f_1 was found experimentally to be equal to that of f_3. The pressures in the ear were then calculated, and it was found that for any two frequencies f_1 and f_2 the results could best be expresssed in the form

$$\log_{10} P_{f1} = A + B \log_{10} P_{f2}$$

where P_{f1} is the root mean square pressure of the tone f_1 at the tympanum when f_1 was judged equal in loudness to f_2, its pressure being P_{f2}, and A and B are constants for the two frequencies f_1 and f_2—A representing the relative sensitivity of the ear for these two frequencies and being dependent only on the pressures at the threshold of audition of the two tones, i.e., on the absolute limens, while $\frac{1}{B}$ is the relative sensibility, i.e., the relative differential limen, at the two frequencies and is constant for the range of frequencies from 100 to 4000 d.v./sec.

Kingsbury used a somewhat similar method, but allowed each of his 22 subjects to vary the intensity himself and to listen to each tone for as long as he wished when making the comparison. He found that, if the amplitudes of pure tones are increased in equal ratios, the loudness of low frequency tones increases much more rapidly than that of high frequency tones, but that for frequencies above 700 d.v./sec. the rate is nearly uniform.

---

[9] Practically all the more recent work has been made possible owing to the development of the triode valve.

IV. Beats, Combination Tones, Consonance, and Dissonance

When two forks or other sources of sound are vibrating with slightly different frequencies, "beats" are heard. They appear as fluctuations in loudness, and occur with a frequency equal to the difference between the frequencies of the sources.

This fact is made use of when the absolute frequency of any sound is required—the number of beats per second which occur when the source is sounded at the same time as a standard source gives the number of vibrations which the source makes in excess or deficit of the standard. If the two tones are so close together that it is impossible to tell which is the higher, the frequency of one of the two is lowered, e.g., by weighting the prong of a fork, and the effect on the number of beats is observed.

Slow beats are by no means unpleasant and, when produced by tones of low pitch, have a solemn effect peculiarly appropriate to sacred music; quick beats are similarly often used to portray agitation or passion. Slow beats especially if produced by complex tones, such as telephone diaphragms, will also be observed to vary in pitch. They are higher in the quieter periods. At such times the fundamentals most nearly neutralize one another, so that their masking effect is less strong, and in consequence the pitch appears higher,[10] due to the relatively more powerful overtones.

Myers (35 p. 37) describes four stages as occurring in the phenomena of beats produced by tones of the middle range. Starting with two tones of 256 d.v./sec. and gradually increasing the frequency of one there is

a) surging, when the beats do not exceed about four or six per second. As the frequency of the beats is increased, the experience changes from one of a regular and rather pleasant rise and fall in intensity to a rapid and discontinuous change, which may be described as a

b) thrusting phase. This phase is particularly noticeable when the beats are about 16 per second. If the tones are sufficiently intense, so that the differences between the maxima and minima can adequately affect the tympanic membrane, the beats arouse tactile sensations, which may be exceedingly unpleasant. As the beats become still more rapid they take on a

c) rattling character, which is most marked when there are about 30 beats per second. These beats are jarring and unpleasant and noisy. With a further increase in their frequency, the "beats" become less perceptible as beats and gradually change to

---

[10] Gueroult (Appendix XIV, Helmholtz) shows that, owing to the physical characteristics of the ear, the pitch, during the maximum of intensity, will lie between those of the primaries, while during the minimum intensity the pitch will be *higher* than either of the primaries if the stronger tone is the higher of the primaries, or *lower* than either if the stronger tone is lower.

d) noisiness, which disappears as the frequencies of the primaries become too widely separated.

Beats from tones of low pitch disappear before the roughness stage is reached, while with high frequencies the rattle "becomes a chirp."

Intertone. If two tones sounding together are producing beats, and one of the tones is stopped, a change in pitch is observed. The pitch of the beating tone, as experiment easily shows, lies between that of the primaries. This intermediate tone is called the "intertone." It carries the beat. Eberhardt (12) reports that with primaries giving eight beats per second, two intertones were to be observed when the intensities of the primaries were equal, or approximately equal. One of these tones was near the pitch of the louder and higher primary, the other was three to five vibrations above the lower primary.

Except when the primaries are very close together they can be heard together with the intertone.

Combination Tones. If two similar tones of the same pitch, e.g., 500 d.v./sec. are sounding, beats will be produced if the frequency of one is increased. As the frequency of the beats is increased and while the beats and tactile sensations in the meatus are still to be perceived, a tone of low pitch gradually creeps in and rises in pitch as the frequency of the higher primary is increased. This tone is the first difference tone, and its frequency is equal to the difference between the frequencies of the two primaries. On this account this tone is sometimes spoken of as a "beat" tone,[11] and the fact that a resonator will pick out the beats and resonate to them has been adduced as evidence for considering the difference tone as a beat tone. But observation shows that both beats and the difference tone may occur together, so that a description which implies that beats and difference tones are due to the stimulation of the same sensory nerve-endings is not correct. Beats and difference tones are, in all probability, due to the stimulation of different parts of the basilar membrane. Their formation is simply explained on a resonance theory of hearing.

Besides the first difference tone, of frequency $(m-n)$, when m and n are the frequencies of the two primaries, various other tones with frequencies equal to other combinations of the two primaries may be heard, e.g., $(m+n)$; $(m+2n)$, etc., and $(2n-m)$, etc. These tones are classed together as combination tones; those formed by adding together the frequencies of the primaries (or multiples of these) are called summation tones in contradistinction to the difference tones whose frequencies are given by finding the difference between the frequencies of the primaries (or of multiples of the primaries).

The first difference tone is produced both externally in the air, where it may be reinforced by resonance, and in the ear, where the other combination tones are generally supposed to be introduced, owing to the asymmetry of the conducting parts. Since cases are recorded of individuals who perceived combination tones after loss of the drum of the middle-ear apparatus (45), it would appear that they may arise in the inner ear itself. This, from a physical point of view, would be expected.

---

[11] This view was put forward by Young in 1800.

Combination tones were at one time supposed to be most easily produced by strong primaries. The mathematical theory follows on the assumption that the displacement of the vibrating mechanism is not strictly proportional to the force causing the vibration, and it was presumed that this would occur principally for large vibrations. But the vibrating mechanism is so asymmetrical that the primary assumption holds even for tones of weak intensity, which should be used for observation. The best results are obtained when the primaries are equally intense.

Beats and combination tones are objective in the sense that they are produced before the excitation of the sensory nerves. They can be produced only by the waves from the two primaries reaching the same ear, and are not obtained when the primary tones are led to the two ears separately (6).

Bone Conduction. We have seen that air-borne sound waves, if sufficiently intense, may set up vibrations in the tympanic membrane which will cause a tactual sensation. It is also well known that the foot of a vibrating tuning fork applied to the mastoid process, or better still to the teeth, can be heard distinctly, the vibrations being transmitted through the bony framework of the head to the cochlea. In this way beats and combination tones, very definitely localized in the ears, are produced when a suitable tone is led to the ear through a tube and the cochlea is stimulated by another tone by bone conduction.

These facts, and some experiments performed by Mader (31) in 1900, have raised the question whether sound waves reaching one ear in the normal manner by air transmission may not be conducted through the head to the other ear. Mader's experiments are not convincing. He experimented on two human heads which had been preserved in formalin, one of which had the top of the skull and brain removed. A specially constructed microphone, placed on various parts of the bones of the heads, indicated that the bones were set into vibration when sounds were led to the ears. Though the microphone was set into vibration, one hesitates to accept the inference that the vibrations were communicated through the bones of the head. Nikiforowsky (37) has been unable to obtain any evidence for such bone transmission, and many other experiments designed to test the assumption have failed to produce any evidence in its support (6).

It appears that, though transmission of sound vibrations through the head is not impossible, the amount of such transmission, when air-borne waves fall on one ear, is exceedingly small. Fletcher (15) found that it was necessary to increase the stimulus at one ear 1,000,000 times over its threshold value before the minimum threshold value for a tone at the other ear was affected.

Binaural Beats. When tones slightly out of tune are led to each ear separately, the phenomena are very different from the beats observed when the tones are led to the same ear or excite the same cochlea directly. A periodic change, the period being the beat frequency, is observed, and to this phenomenon the term binaural beat is applied.

As with the monaural beats, one tone is heard which may change slightly in pitch, but in place of the waxing and waning of intensity there is a peculiar swinging of the apparent source of the sound from one side to the other, at a rate equal to the difference in the frequency of the primaries. The amount of the swing depends on the frequency of the primaries. With primaries of low frequency the movement is from one ear to the other with a lingering at the ears; with increase of frequency the lingering at the ears decreases; and with higher frequencies the movement becomes more and more restricted till, at frequencies variously estimated at from 2000 to 4000 d.v./sec., it is so minute that it cannot be discriminated.

This phenomenon depends on the fact that the waves arriving at the two ears stimulate what may be called "corresponding points" and give rise to the perception of a single tone. This tone appears to originate from a single source. At the same time, since the frequency of one tone is slightly greater than that of the other, the waves stimulating that ear arrive at an earlier time, as compared with the waves which stimulate the other ear, till at last they are one whole wave-length in advance. This is exactly the effect which would be obtained if a source of sound were moved from the median plane so far to one side that the difference in the lengths of the paths of the sound to the two ears were half a wave-length, and if then the source were suddenly at an equal distance on the other side and gradually moved to the median position again. The sensations are, in fact, so interpreted. The locus of the movement depends on the individual. Some subjects report that the locus is semicircular or elliptical in front; others that it is behind or over the head. Also it may be external or internal to the head. These individual differences are probably secondary and not due directly to the auditory sensations.

These phenomena were not observed by the earliest investigators, probably because the two tones were not sufficiently isolated, with the consequence that each ear was stimulated by both tones, and monaural beats were produced in the ordinary way. Both S. P. Thompson and Rayleigh observed the beats as the apparent movement of a source of sound, and this has since been fully substantiated (53).

If the phenomenon gives rise to the judgment that the source moves nearer or farther away, a corresponding change in loudness is reported —quieter at the greater, and louder at the nearer distance.

After-Sensations. As the auditory sensation takes time to develop, so the sensation persists after the stimulus ceases.

The amount of the persistence has been estimated in various ways. Helmholtz took as a measure the rate at which two notes could be alternated in a trill, without fusion of the two consecutive sensations into one sensation. Abraham used the same method, obtaining the trill from a siren with two rows of holes producing different tones and blocking such holes in the two rows that the two tones were sounded alternately. Others (32, 52, 57) have used a "flicker" method, the tone being interrupted a certain number of times a second, and the persistence obtained from the maximum time of interruption to produce a continuous sound. But all these methods must give persistence times greater than the true persistence times,[12] since they include both the waxing and the waning. Bishop (8) experimented by cutting out the sound altogether and calling for introspective reports. The experiments are very difficult owing to the impossibility of insuring that vibrations are not communicated by the source after it is supposed to have been cut out. He found no evidence for a positive after-sensation, i.e., for a continuance of the tone, after the stimulus was removed, but "modified endings" were reported. These he ascribes chiefly to apparatus noises, but of this there is no proof. It may be that the sudden cessation of the stimulus produces vibrations in other parts of the basilar membrane, causing a sensation of noise as would be expected on a resonance theory of hearing, so that the modified ending may be, in part, sensory.

There is no evidence for a negative after-sensation.

Consonance and Dissonance

The musical scale. In music the various tones are named according to the first seven letters of the alphabet. The middle note of the piano is called middle C. For scientific purposes this tone, which is usually written c^1, has a frequency of 256 d.v./sec.; the frequency of c^1 used in music has a slightly higher frequency, about 270 d.v./sec.

When two tones, the frequency of one being variable, are sounded successively or together, it is found that certain frequencies fuse much more than others. These frequencies give a unitary impression when sounded together, and when one only is sounding it may be confused with the other even by practiced musicians. The tones most frequently confused are those whose frequencies are in the relation of 1:2. Such tones are said to be the octaves of each other. Thus the octaves of middle C (256) are the octave above, c^2, with a frequency of 512 d.v./sec. and the octave below, c^0, with a frequency of 128 d.v./sec. The octaves higher still, c^3, c^4, etc., have frequencies of 1024, 2048 d.v./sec., etc., while those below, c_0, c_1, have frequencies of 64, 32 d.v./sec., etc.

After the octaves, whose frequencies are in the ratio 1:2, tones are found to fuse together when their frequencies have the ratios 2:3; 3:4; 4:5; 3:5; 5:6. But frequencies with these ratios do not fuse so readily as do the octaves 1:2.

The frequencies, relative to c^1, which fuse most easily are, therefore,

$\frac{3}{2}$ of 256 = 384 d.v./sec.—the frequency of the fifth above middle C

$\frac{4}{3}$ of 256 = 341.3 d.v./sec.—the frequency of the fourth above middle C

$\frac{5}{4}$ of 256 = 320 dv./sec.—the frequency of the major third above middle C

$\frac{5}{3}$ of 256 = 426.6 d.v./sec.—the frequency of the sixth above middle C

---

[12] Kucharski thinks that any persistence there may be must be extremely small.

$\frac{6}{5}$ of 256 = 307.2 d.v./sec.—the frequency of the minor third above middle C

It is from these frequencies that the musical scale is constructed, which is

c^1	d^1	e^1	f^1	g^1	a^1	b^1	c^2
256	288	320	341.3	384	426.6	480	512

It will be seen that all the frequencies calculated above, except that of the minor third, fit into the above scale. The minor third falls between the second and the major third.

It is possible to calculate the frequencies of a scale beginning with any of these frequencies. For example, the fifth and third of f^1, which should give c^2 and a^1, respectively, are

$$\frac{3}{2} \text{ of } 341.3 = 512 \text{ d.v./sec.}$$

$$\frac{6}{5} \text{ of } 341.3 = 409.5 \text{ d.v./sec.}$$

Though the fifth gives exact agreement, the third does not—we have a new frequency 409.5 for a^1 whereas we need a frequency of 426.6 for a^1 when it is played in the scale of c^1. And so it is found that an enormous number of tones are required if music is to be played in tune in every possible key. To overcome this difficulty, the octave is divided for practical purposes into twelve equal intervals called semitones, the equality of the interval being dependent on the *ratio* of the frequencies, not on the numerical difference in the frequencies.[13]

Theories of consonance. When two tones are sounded, each will have its own overtones, while the two together will produce combination tones. Any two of these numerous tones will produce beats if their frequencies are sufficiently close together. If these beats are sufficiently powerful they will give a roughness, cutting the tones up into separate pulses, which will be decidedly unpleasant. This, according to Helmholtz (22), is the cause of dissonance. Consonance, on the other hand, is a continuous even flow of tone, undisturbed by beats, the tones being also bound to each other owing to their having coincident partials.

Stumpf (50, Sec. 19, 20) considers this explanation is not sufficient. There are many dissonant intervals, he maintains, whose partial and combination tones are too far apart to produce beats, e.g., 700:900, 800:1100, while many intervals, e.g., the third in the lower octaves, remain consonant although the frequencies of the two tones are sufficiently near to one another to produce beats. He considers tonal fusion to be a psychologically elemental unity, incapable of further analysis; but he does not explain why certain frequencies fuse and others do not.

The changes which music has undergone in recent years, and the very different ideas of consonance and dissonance held by older and younger musicians—the former perceiving only dissonance where the latter find harmony—indicate the possibility of consonance's being built up during the course of the individual's life. Harmonics, etc., are consonant probably because a body vibrating freely vibrates with frequencies which are multiples of the fundamental, so that these tones, most frequently heard together, are the ones which are regarded as fusing.

---

[13] It is for this reason that in almost all the diagrams and in the text, equal distances represent equal *ratios* of frequencies, not equal *numerical* differences.

V. Auditory Perception of Space

Auditory Space. Into the auditory perception of space, many and various extra-auditory factors enter. For several reasons our unaided ears are ill qualified to perceive space so that when a sound is heard it is not with our ears alone that we endeavor to localize the source; perhaps even more we search with our eyes. Into the whole process, there enters a tendency to turn towards the source when localizing the position; even when no movement occurs, the muscular set is probably there.

The importance of the eye is well illustrated in some experiments of Young (59). He had ear-trumpets fitted in front of his ears, the ear-trumpet on the right side leading to the left ear, the trumpet on the left to the right ear. In consequence, his auditory cues were reversed—a sound on the right appeared to come from his left, e.g., a horse trotting down the street from left to right was heard as going from right to left. But, and this is the important point, if he saw the source of the sound, the visual cues were dominant. When the horse was seen as well as heard, the visual cues dominated the situation; the horse was perceived as going in the proper direction—left to right. At the end of 18 days, during the last three of which the trumpets were worn continuously (except when in bed, when the ears were plugged), sources of sound were correctly localized if the eyes were used; yet, with eyes closed, every trial gave right-left reversal. So that, though our localization of a source is not solely auditory, our ears furnish the clues. It is the nature of these auditory clues which will be considered in what follows.

One of the many reasons why our auditory perception of space is so inaccurate is the fact that sound waves, unlike light waves, do not travel in a straight line. They are deflected from the straight-line course by the wind, and even in still air the direction of travel is changed owing to temperature variations. They bend round obstacles by diffraction and suffer reflection at so many surfaces that the direction from which the sound waves which excite the ears actually come may be totally different from the true direction of the source. This is well illustrated in the case of echoes: we hear first of all the sound waves from the source near us and after a time we hear the sound waves coming from another direction altogether after reflection by some building or mountain, though we know the source is by our side. The wonder is that, with the immense number of trains of waves which under ordinary conditions reach our ears from various directions, all emanating from one source, we are able to perceive with any accuracy whatsoever the true direction of the source. In open country the primary wave travels in a more or less direct line from the source and is much the most dominant wave, so that the localization of the source of the sound is more accurate; yet the localization of the source, even by wild animals, is not absolutely accurate, though it is sufficiently so in general. The animals are enabled to orient themselves towards the direction of the source for which they search with their eyes and noses.

Besides these physical difficulties, perception of the direction of a novel sound source is further handicapped by the fact that the single ear has no means of discriminating differences of direction between approaching waves. It has nothing which corresponds to "local sign." In this the ear differs from the eye. In the eye, images of objects in different directions are formed at different points on the retina, the angular distances between the retinal images being strictly dependent on the angular distances separating the objects. Yet, in spite of this fact, an individual totally deaf in one ear can often judge with fair accuracy both the direction and distance of well-known sound sources, either because he recognizes the source and knows where it is, as, for example, a motor passing his house in the street outside; or because he recognizes differences in the quality, loudness, or other attributes according to the position of the source, as, for example, the variations in the loudness of a pistol shot with distance.

But, ruling out all such "complications" which are perhaps the most important in ordinary life, the single ear is no better qualified to tell the direction of an unknown source of sound than the single eye is to estimate the relative distances of two unknown objects situated alone in space. With two ears we can, however, tell direction within certain well-defined limits; to perceive the distance of an unknown and unfamiliar source of sound a third ear would be necessary—in normal life familiarity enables us to judge distance approximately owing to the "complications."

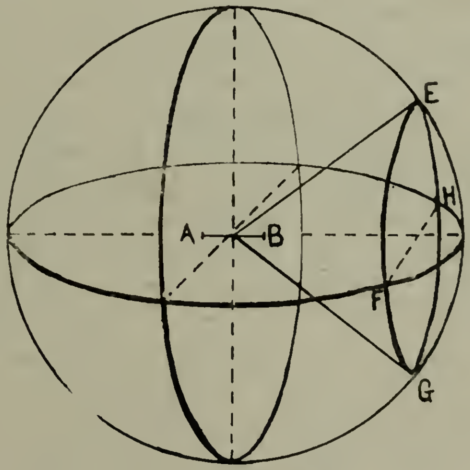


FIGURE 13

The limits just mentioned of our powers of perceiving direction are due to the fact that it is the appreciation in some manner, which will be discussed below, of differences between the wave motions at the two ears which furnish the data for the perception of direction, and it is evident that the differences produced at the points *A* and *B*, which we may take to be the two ears, by a source at *E* would also be produced by the source if it were at any other point on the circle *E F G H*—in other words, the ears can give direction only on the surface of a cone; on the information as given by the binaural differences there can be no discrimination between a source at *E* and one at *G*. But here "complications" help—we are accustomed to deal chiefly with sources on the earth, hence, if the source is known to be external, it is usually located on the ground level, seldom in the air unless it is recognized as coming from a bird or from an aeroplane. Deer and other wild animals practically never look up into trees, and a dog, brought up in the country, if spoken to when outside, from an upper window, will not look up; he will look in the right direction, but at his own level. In experiments on "sound" localization, when the ears are stimulated separately, great variations in actual localization are found between individuals: some maintain that the phenomenal effect is that of a source moving round the head in a horizontal plane behind them, others report that it is as though the source moves in a vertical plane, while for others the source seems to move in an intermediate position. All these positions are equally correct interpretations of the auditory sensations. It is an interesting fact that most observers localize the apparent source behind the line joining the ears, and often, especially when telephones are used, inside or very close to the cranium, while blind observers state that it seems as though the source were in front of them—the apparent source is practically never below the level of the ears.[14]

Realizing the numerous factors which, in ordinary life, enter into our auditory perceptions of direction, what are the important factors in sound waves for such perception? In what does the wave motion at one ear vary from the wave motion at the other ear? It must be some differential effect which enables us to perceive the direction from which the sound waves approach.

The sound waves emanating from any given source may differ at one ear from those at the other in many ways, viz., in (*a*) amplitude, the vibrations being greater at one ear than at the other, so that there are binaural differences of loudness; in (*b*) form, for, as we have seen, the head is a greater obstacle to waves of short wave-length than to waves of long wave-length. The result will be that the vibrations at the far ear will be simpler than those at the near ear if the vibrations are complex, for at the latter ear the shorter wave-lengths will be more prominent; in (*c*) phase, for the vibrations will be "in phase" if the source is in the sagittal plane (the plane at right angles to the line joining the ears), while there will be a difference of phase at the two ears as the source is to the right or left of this plane; and in (*d*) the time of arrival of the same part of the wave at the two ears. This is not the same as (*c*) (phase), for, if two waves of different frequency, and therefore of different wave-lengths, start at the same time from

---

[14] I have only once out of thousands of observations had a report from a normal subject that the source appeared to be lower than the line joining the ears. This was when an observer noticed a tuning fork on the ground. Blind observers will say, "It is as though the sound comes from a point there on the table in front of me."

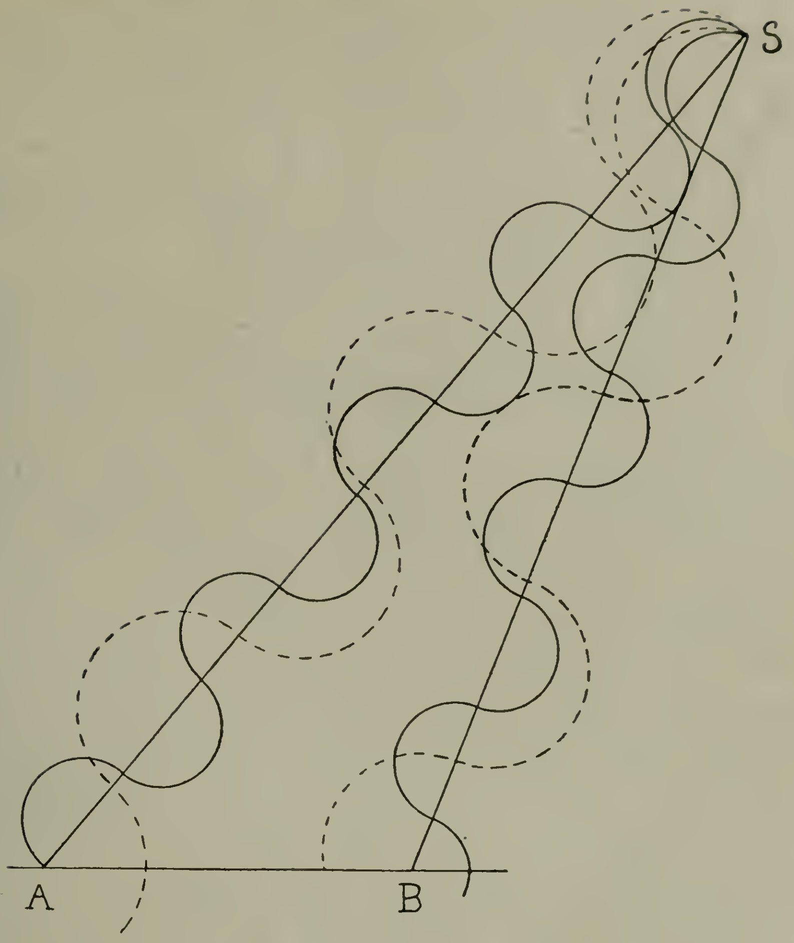


FIGURE 14
Two Sine Curves of Different Sizes So that the Two are in Different Position at *A* and *B*

a source (*S*), Figure 14, the difference between the times of the arrival of the same given waves at *A* and at *B* will be the same for both frequencies, since the velocity of sound in air is independent of frequency, but the difference in phase at *A* and *B* for the longer wave will not be the same as the phase difference for the shorter—the difference of phase of the shorter wave will be less than that of the longer in the circumstances shown in the figure.

All these factors have, at one time or another, been put forward as being at the basis of our ability to localize, and it seems probable that they all, supplemented by the various "complications" which have previously been noted, assist, with the possible exception of "phase," which was at one time strongly championed.

Intensity differences were at first considered to be most important, though Rayleigh pointed out that for sounds of low and medium pitch the head offered no impediment so that for such sounds intensity differences would not exist. At the same time, since practically no sound is pure, it was always open to argument that, because the head was a considerable obstacle to waves of high frequency and short wave-length, the higher components were shut off at the farther ear and that this resulted in intensity differences, and that, at any rate, there would be differences in timbre. That this was not a satisfactory explanation, Rayleigh himself showed in an experiment[15] by means of which he altered the relative phases of the same sound at the two ears keeping the intensity and the timbre at the two ears constant. He demonstrated that the apparent position of the source varied, as the waves arrived at one ear before the other, in spite of there being no change otherwise.

Myers and Wilson (58) then put forward their hypothesis that, for low frequencies, direction was perceived on the basis of the phase difference at the ears. Assuming that when one ear is stimulated by a sound wave the excitation is transmitted through the bones of the head to the other ear, and making further assumptions which need not detain us here, Myers and Wilson showed that this difference of phase resulted in a difference of intensity at the two ears, the intensity being greatest at the ear nearest the sound source. The hypothesis receives a certain amount of experimental support, for many observers give the same judgment of direction for the same difference of phase over a range of frequencies. But we have to remember that we all have position habits; that accurate perception of direction is not necessary in ordinary life—a general idea of the direction accompanied by searching with the eye being more useful; and that the direction is not constant for a given phase difference. The displacement to one side depends on the frequency, being less for high than for low frequencies, if the phase difference is constant. The whole hypothesis has recently been subject to very close scrutiny, which has not resulted in a finding in its favor (3).

The time-difference hypothesis was later put forward on the basis at first of experiments with "clicks," i.e., of sound waves, complex in form and of very short duration. It has since been extended to the explanation of the perception of the direction of waves from a continuously vibrating source.

According to this hypothesis (4) an appreciation of the interval of time between the arrival of the same sound wave at the ears furnishes the localization cues. If the source is in the median plane, the successive waves arrive at the ears synchronously; if the source is to one side, the waves reach one ear before the other, the maximum interval of time being that necessary for sound waves to travel from one ear to the other.

It is suggested that when a sound wave falls on the tympanum, forcing it in and out, a certain group of fibers of the basilar membrane, depending on the frequency of the exciting tone, vibrates up and down. If the sound source is in the median plane the corresponding groups of fibers in the two cochleae will vibrate together; if the sound source is to the right, the fibers in the right cochlea will be slightly ahead of those in the left ear by an amount depending wholly on the difference between the distances of the source from the left and the right ears.

Now it is improbable that the nervous impulses are initiated during the upward movement of the basilar membrane fibers. They are initiated when the fibers are moving downwards and the cilia of the hair-cells are under tension. The impulses will therefore be shot off in pairs from the two cochleae, the impulses from the right cochlea leading, if the source is on the right. An appreciation of this time difference, Banister suggests, gives the clue for localization. What happens when the impulses reach the cortex, and how the temporal difference comes to be interpreted as a difference in direction, must for long be a matter of pure conjecture. On the assumption that the cochlea functions in this way, and taking account of the "all-or-none" nature of the nerve response, an explanation is offered of all the known facts of localization, from the apparent localization of many sources when the pitch is high to the ambiguity of localization which occurs when the time difference exceeds a certain maximum. This time difference corresponds approximately to the distance round the head.

---

[15] Rayleigh (41) used an apparatus for varying the phase which has been elaborated by Stewart (48) and used extensively in his investigation.

On this hypothesis, too, it is obvious why complex sounds are so much more easily localized than purer tones. With complex sounds, the same cues are being furnished by large numbers of fibers; with pure tones but few cues are available.

Volume. Sounds have volume. Some are small, compact, and point-like; others are big and voluminous and occupy much space. Tones of high pitch are small; tones of low pitch are large.

Volume, according to Watt (55), is one of the elementary attributes of sound. He associated it with the wave-length of the sound. Physiologically, volume, he said, corresponded with the length of basilar membrane affected by the tone which, in turn, corresponded to the area of the strip of the cortex excited. Against this hypothesis it may be urged that it fails to explain why the discrimination of change of volume is coarse as compared with that of change of pitch. The pitch of a tone is determined by the position of maximum disturbance of the basilar membrane, which is, Watt states, at the middle of the length affected by the tone. From this it follows that for unit change of pitch there will be a change in volume in which the one dimension varies by two units, and this shortening is *always* from the broader end of the basilar membrane since the higher tones are heard through the vibration of the end of the basilar membrane which is nearest to the fenestra ovalis. Therefore, any change in pitch is *always* accompanied by a much *larger* proportional change in volume. How is it then that the large changes are so difficult to follow? And how, on the other hand, is it that a change in pitch, which is solely a change in the prominences or ornamentations of the outline of the volume, a very slight shift of an apex to the right or left, is appreciated without the much larger change in the whole being sensed?

One can but wonder what are the functions of the nerve-endings and of all the complicated structures of the upper half of the basilar membrane if they are to serve neither for pitch discrimination nor for volume. For the lowest pitches, one would expect volume changes to be comparatively inappreciable and pitch discrimination to be very fine. Yet here, Watt states, the reverse is true: volume changes can be appreciated, not pitch changes (55, p. 29).

But all this might be explained were it not that the volume of a given tone is not constant—not only does it vary from individual to individual, and in the same individual from day to day, but also it varies in the same individual from moment to moment, while the pitch remains, save in exceptional cases, [16] constant. Since the pitch of a pure tone is determined, according to Watt, by the middle point of the volume, it must be impossible to have a change in volume without a change in pitch. Yet such changes occur continually. May it not be, as has been suggested (2), that the description of a sound as having volume is due to the fact that, for some reason or other, the "direction" from which the sound waves reach the ears is not definite? The suggestion made is that, if the direction is clear and unequivocal, the "sound" is judged to be small; if it is indefinite, its volume is judged to be great; it will vary with the amount of the indefiniteness. If this is so, it is to be expected that sounds of low pitch will be more voluminous than sounds of high pitch, for the former, owing to their long wave-length, bend around obstacles and are reflected to the ears from all directions, whereas sounds of high pitch travel much more definitely in straight lines.

From our consideration of the auditory localization of sources of sound, we have seen that the chief factor in such localization, when the binaural intensities are constant, is the difference of time between the arrival of the waves at the two ears. It is obvious that for a given angular movement of the source (*S*), Figure 15, the time difference will be a maximum in a direction *XX*, at right angles to the line joining the two ears, and a minimum in the direction *YY* in the line of the ears, *AB*.

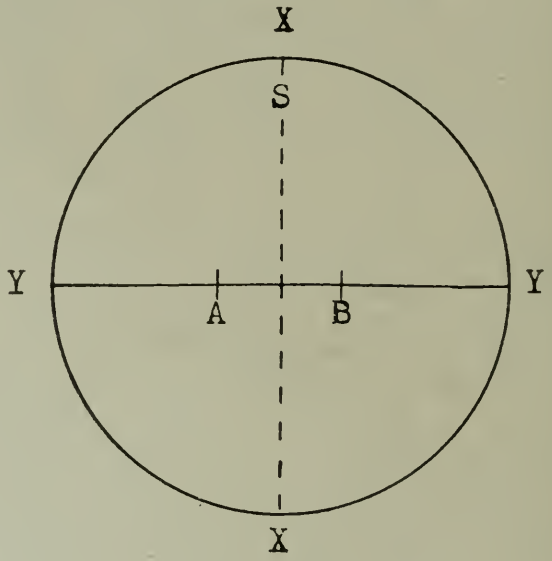


FIGURE 15

So we would expect to find that a "sound" localized in the median plane had a greater volume than one localized in the aural axis, for a slight movement, e.g., of the head, causes a much greater change in the apparent direction of the source in the median plane—a well-known

---

[16] Where there are binaural differences of pitch (diplacusis).

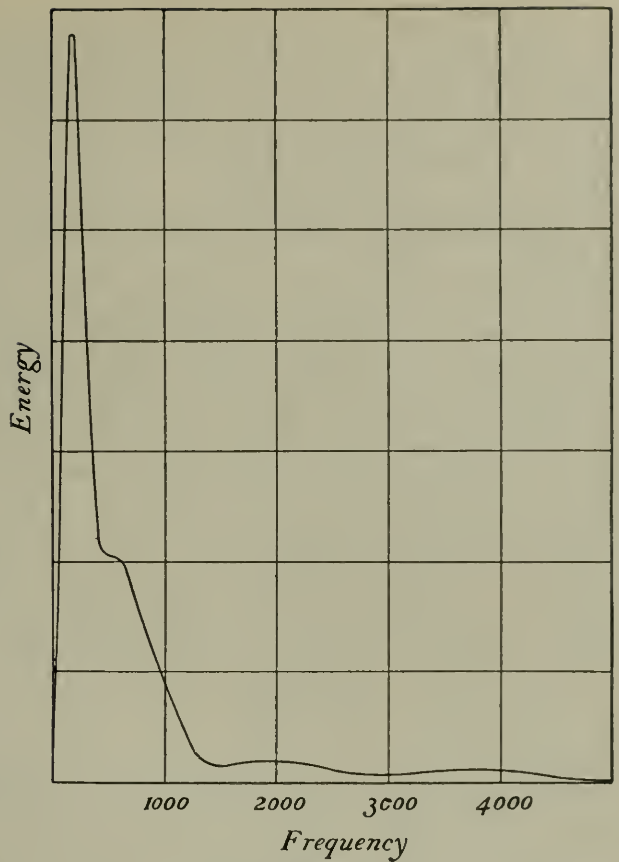


FIGURE 16
Energy Distribution over Various Frequencies in Normal Speech

experimental fact. Young (59), too, in his experiments with transposed hearing, found that volume seemed to change with precision of localization. Sometimes a sound was heard precisely at a single spot. At other times, the same sound was larger and was visualized as covering an area. A sound, vaguely "off at the right," appeared more diffuse than when definitely localized. Most interesting of all were the unlocalized sounds, in which the pitch, intensity, and timbre were about as usual but the sounds had no particular place or character; they sometimes surrounded the body diffusely. A leak artificially made around the ear-pieces was found to make localization indefinite and to give the sounds greater "outspread."

Thus, we see that the evidence is weighted against "volume" being a fundamental attribute of sounds; but there is need for further research before this pronouncement is finally passed.

VI. Speech and Deafness

Speech. Speech employs frequencies from about 75 to above 5000 d.v./sec., with a great concentration of energy in the region below 1000 d.v./sec. (11). Yet, in spite of the comparatively small amount of energy carried by frequencies over 1000 d.v./sec., these frequencies appear (14) to be most essential, if the speech is to be recognized, as is shown when the higher frequencies are eliminated. For example, if all frequencies below 500 d.v./sec. are eliminated, the decrease in the comprehensibility of the spoken syllables is said to be reduced by only 2% though the intensity is reduced by 60%; on the other hand, if all the frequencies above 1500 d.v./sec. are eliminated, comprehensibility is reduced by 35% though intensity is still 90% of the original intensity. In particular, the sounds *s, z,* and *th* are very definitely affected when frequencies above 5000 d.v./sec. are eliminated.

There is not as yet universal agreement even as to the nature of vowel sounds, which consist of simpler combinations of tones than consonantal sounds, and which are therefore more easily studied. According to one theory, the higher tones are overtones of the fundamental, these overtones being reinforced by resonance in the cavities of the mouth and head. According to another theory, the high tones are fixed for the various vowels and are independent of the frequency of the fundamental, being decided by the natural frequencies of the oral cavities.

The optimal intensity Fletcher found to be about 1/1000th that at a distance of half an inch in front of the mouth of the speaker. It was not till the intensity was decreased to 1/1,000,000th that speech began to show any loss of clearness; at low intensities consonants are, as a rule, more difficult to follow than vowels.

Deafness. The phenomena of deafness are well known. Besides the difficulty in hearing, those suffering from any considerable hearing disability often have a deaf voice, owing to lack of proper resonance in the naso-pharynx and nares, and a deaf look, due more to the necessity for keeping the mouth open to breathe than to their inability to follow the conversation.

Inner-ear deafness is not in general very common. It is, however, very frequently found among boiler-makers. A recent inquiry indicates that in Leningrad (44), 50% of boiler-makers suffer from inner-ear deafness after 10, and 80% after 20 years' service. In persons of other occupations, deafness is most usually due to some affection of the middle ear, e.g., catarrh of the middle ear and obstruction of the Eustachian tubes with or without suppuration; or to otosclerosis, with fixation of the foot of the stapes in the fenestra rotunda. It may also be due to injury to the tympanic membrane, though a simple perforation has little effect on hearing.

Obstruction of the Eustachian tubes is the commonest cause of deafness. For efficient hearing, the pressure on both sides of the ear-drum must be equal. Otherwise the membrane is not free to respond to the pressures of the sound waves. This equality is maintained normally by the opening of the Eustachian tube every time we swallow. If the Eustachian tubes are not blocked, swallowing is accompanied by a click in the ears. If the Eustachian tube is blocked, suppuration in the middle ear will cause the tympanic membrane to be forced outwards; if there is no suppuration, the air pressure on the inner side will be decreased, and the membrane will be displaced inwards. In either case there will be loss of hearing.

Treatment in such cases aims at doing away with the infection and equalizing the pressures on the two sides of the ear-drum. The latter operation usually consists in blowing air through the Eustachian tube into the middle ear with a catheter; but this does not often effect a permanent cure. It alleviates the symptoms temporarily, but forces the mucous, etc., with which the tube is blocked into the middle ear, from which its only means of escape is via the Eustachian tube. The result is that the tube is quickly blocked again and deafness recurs. Nesfield (36) suggests opening a small passage through the mastoid process to connect the middle ear with the external auditory meatus. In this way he claims to have cured many cases of deafness.

The usual clinical tests for acuity are not very accurate. Common tests include a determination of the distance at which a watch or the voice can be heard. More accurate results are to be obtained with a standardized source of sound, such as Politzer's acoumeter, a small hammer which falls onto a metal bar from a standard height.

Such examination serves the purpose of roughly grading the amount of deafness, but does little more, for an individual may be deaf to a low-pitched voice and yet hear other sounds comparatively easily. For accurate estimation it is advisable to test acuity, for example with the audiometer, for many tones throughout the tonal range.

Experiment has shown that, whereas the normal ear requires a pressure of 1/1000 dyne/sq. cm. for hearing, those who require 1 dyne/sq. cm. pressure can still carry on though they are noticeably deaf, while the very deaf, those who use trumpets and other instruments for the deaf, require a pressure of 10 dynes/sq. cm. (17). Incipient deafness usually affects the limits of hearing initially. There may be almost complete loss of hearing for both high and low tones, combined with comparatively good hearing for the frequencies most usually employed in speech, say from about 200 to 2500 d.v./sec. Hence incipient defects should be sought outside these limits. Contrary to what might be expected, the differential limens for loudness, as well as for pitch, appear to be the same for the abnormal as for the normal ear, provided the intensity of the sound is sufficient for the former to hear.

Paracusis. Experiments on the oft-alleged ability of many deaf people to hear better in a noise (paracusis) show that, in spite of the deaf person himself believing he hears better in a noise, he does not; his hearing becomes more and more interfered with as the loudness of the noise is increased (24, 40). Yet, in certain cases, as Knudsen and Jones point out, he may have a relative advantage. The paracusic suffers from a conductive impairment, which especially affects the hearing of tones of low pitch (below 512 d.v./sec.); for tones of higher frequency his hearing is often fairly acute. Now most noises commonly experienced (the noise of a train and of traffic) are relatively low in pitch, and low-pitched sounds are relatively more efficient in masking higher tones, such as are used in speech. So the paracusic has the advantage since his disability filters off the masking noise which interferes so markedly with the normal ear. He also has one other advantage since in the presence of a noise the tendency is to speak louder. Such increase in intensity produces distortion, which again affects the normal ear but not the paracusic, because the latter's conductive impairment reduces the intensity of the sounds at the cochlea so that the intensity is the normal to him.

VII. FATIGUE

One of the most pressing problems of modern civilization is the effect of noise on the organism.

From the biological point of view, it is perhaps to be expected that the ear should never, even in sleep, suffer any diminution of sensitivity. Unlike the eye, the other principal distance receptor, the human ear has no flap or lid to shut out stimuli at will, and indeed the importance of being awakened immediately on the approach of danger, may have tended to an increase of sensory acuity during sleep combined with a heightened perceptual threshold owing to the mental inhibition of all but certain specific sounds.[17] With the sudden rise of the mechanical age, the problems connected with the effects of noise are being forced more and more to the front. At first, inquiry was confined to certain specific problems such as the consideration of boiler-makers' disease, where loud continuous noises produced physical lesions; now the great concentration in huge modern cities, with the incessant roar of the traffic, has made the question of auditory fatigue one of first importance.

Fatigue is, in general, evidenced by inability to respond to a stimulus, and whenever there has been any failure of response to a stimulus normally effective it has, far too frequently, been ascribed to fatigue. This is illogical, unless the term "fatigue" is used in so wide a sense as to make it valueless. Failure to respond may be due to the fact that attention is directed elsewhere, as when one is so absorbed in a book that the dinner bell, however loudly it may be rung, is not heard; or failure to respond may be due to adaptation, as when a continuous stimulus is neglected and we can respond to other stimuli as though the one to which we are adapted no longer acts. This is obviously not an instance of "auditory" fatigue, though it may result in a state of general fatigue. We shall return to this later.

If the term auditory fatigue is to have any significance, it must be confined to effects, produced in the auditory sensory apparatus by previous stimulation, in consequence of which there is a diminished ability to respond. It is in this sense alone that the term auditory fatigue should be used; loss of acuity, for example, should rigidly be ascribed to auditory fatigue when, and only when, it is maintained that the loss is due to some effect, some change, brought about in the sensory apparatus by the fatiguing stimulus.

All sensory fatigue appears to be due directly or indirectly to metabolic chemical change. The direct effects are similar to those which would be involved in the eye if Hering's hypothesis were true. It would then be necessary to assume that somewhere in the auditory system there was a chemical substance which underwent change while the ear was being stimulated, and that, so long as the total amount of the substance was being changed, the sound was audible—the sound, if continuous, would continue to be heard till the rate of the change produced by the stimulation was equal to the rate at which the change was being neutralized automatically. At this latter stage the ear would be fatigued for that particular sound. If such an auditory substance exists, it is probably located in the hair-cells of the organ of Corti. To account for the fact that sounds produce no negative after-sensations, we would be forced to the further assumption that it was a change in the substance in one direction only which initiated the nervous impulses which are the basis of the sensation. In this, the hypothesis would differ markedly from a similar hypothesis as applied to the other sense-organs. Such fatigue would be specific to the components of the sound which produced the fatigue.

---

[17] One thinks immediately of the response of a sleeping mother to the slightest movement of her sick child.

What has been termed indirect metabolic fatigue may be expected where muscular activity is part of the sensory response. It would correspond to muscular inhibition. When the metabolic change has reached a certain point, inhibitory nervous impulses are initiated which prevent further muscular action. Such fatigue might be expected in the tensor tympani and stapedius muscles, if one assumes that contraction of these muscles is necessary for audition. It would probably be non-specific, being general for all sounds.

Experimental work, which has obviously involved measurements of some form of sensibility or sensitivity both before and after the ear has been subject to the "fatiguing" stimulus, has proceeded along four main lines. These are the measurement of

a) the change in the absolute threshold of a sound, well within the auditory range;

b) the change in the apparent loudness of a sound, the "unfatigued" ear being used for purposes of comparison;

c) the effect on the apparent position of a binaurally localized sound source, of monaural fatigue;

d) the effect on the persistence of auditory sensations.

a) *Absolute threshold.* The absolute threshold before and after "fatiguing" has been determined by various methods by Urbantschitsch (52), Flügel (18), and Achmatov (1).

In one method the minimum current through a telephone which would give an audible sound was measured; in another, the length of time a feebly vibrating tuning fork could be heard was taken as a measure of the threshold; in a third, the unfatigued ear was taken as a standard, the two ears having been made equally sensitive in an unfatigued state.

It is perhaps unjustifiable to make the assumption that a change detected by the tuning-fork method is comparable with one obtained by the other methods, since the psychological processes involved are so diverse. But effects by these methods have been observed which, with many exceptions, are as follows:

(1) previous stimulation causes a heightening of the threshold, i.e.,

the fork is heard for a shorter time, or the telephone diaphragm must vibrate with greater intensity for the sound to be perceived;

(2) the effect is specific to the fatiguing frequency; in one instance it was found to be confined to about one difference limen;

(3) the intensity of the fatiguing sound has little effect on the resulting "fatigue";

There are considerable differences as to

(4) the duration of the "fatiguing" stimulus which produces the maximum effect; some put this at 40 seconds, others at several minutes; and

(5) the length of time the effect lasts; it is said to vary from 30 seconds to over 20 minutes.

b) In the second method, the effect of stimulating one ear for a certain period is found by a direct comparison of the loudness of sounds at the two ears.

This method has been used by Müller (1871), Huijsman (1884), and by Pattie (1927). Huijsman found that, if the tone of a tuning fork (250 d.v.) was led by a bifurcating tube to both ears, the effect of a two minutes' continuous stimulation of the left ear was to make the tone appear louder to the right ear when the tone was presented to the ears successively. These findings have been confirmed and extended by Pattie (38), who used a valve circuit for the production of the tones. Pattie's general method was to keep the intensity in (say) the left telephone constant, and to present the observer with two sounds successively, one in the left telephone and the other in the right. The intensity in the right telephone was varied by changing the current in that telephone. The observer was required to judge which telephone sounded the louder, and the percentage number of judgments of the right being the louder was calculated. The left ear (i.e., the one at which the intensity was constant) was now subjected to continuous uniform stimulation for one minute, after which judgments were again required of the relative loudness of the telephones on the right and on the left, using the same current strengths as before. If the minute's stimulation had any effect it should be evidenced by a change in the percentage number of judgments of the right being the louder.

The general results were

(1) that the minute's stimulation produced an apparent decrease in the loudness of the tone of slightly over one difference limen;

(2) that this apparent decrease of loudness disappeared gradually in from 10 seconds to 60 seconds;

(3) that the effect was not confined to the pitch of the tone used during the period of continuous stimulation, but extended to all tones tested;

(4) that the effect, both in amount and duration, increased with the intensity of the continuous stimulation, though, if the latter were very weak, it acted in the opposite direction, making the ear more sensitive;

(5) that the effect varied with the duration of the continuous stimulation for periods of under two minutes; this was not by any means general. In some cases a two-minute stimulation produced less evidence of fatigue than stimulation for one minute.

c) Experiments on the effect of previous monaural stimulation on binaural localization have been carried out by Sewall (46) and Flügel (18). Sewall found that monaural stimulation for one or two minutes produced a set towards the other side, so that a large difference of intensity at the ears was necessary if the sound, when heard with both ears, was to be localized in the median plane. Such change of localization was, he found, not a gradual change, it was very sudden. The sound appeared at one time to be in one place; at the next moment it was somewhere else. Also he found that an interruption for less than a second caused the whole effect to disappear.

Sewall's method of attempting to change the position of the binaural sound by changing the relative intensities at the ears assumes that localization is an intensity effect, which it is not. Many experiments (5) have shown that a very large change of the intensity ratio can be made without affecting localization in the least. So Sewall's experiments merely show that monaural stimulation causes a set to one side.

This set was, however, measured by Flügel, who compensated for it by varying the time difference at the ears. Flügel found that there were marked individual differences not only between subjects but also between the two ears of the same subject. In general, however, he found that the effect

(1) varied with the duration of the monaural stimulation, reaching a maximum after about 1½ minutes;

(2) lasted for periods up to 12 minutes;

(3) was general; monaural stimulation with one tone caused one-sided localization for other tones heard binaurally;

(4) was independent of the intensity of the monaural stimulus;

(5) was independent of such distractions as reading, looking at pictures, and even of the drowning of the monaural stimulus by a loud tin drum.

d) We have already seen that there are some differences of opinion with regard to persistence—some investigators maintaining there is none. Helmholtz considered there was persistence and suggested that fatigue might increase its duration.

An investigation on these lines for frequencies between 140 and 280 d.v./sec. has been undertaken by Weinberg and Allen (57) who used, as sound source, a Stern Tonvariator enclosed in a box lined with very soft thick felt. In one side of the box was a hole through which the sound issued. Periodic interruptions of the sound were obtained by rotating a light aluminum disc in front of this hole. In the disc were four symmetrical holes, each of the same size as the hole in the box. The disc was driven by an electric motor, the speed of which could be varied. In this way, they measured the maximum length of interruption which just gave the perception of a pure tone and thus obtained a measure of the minimum "persistence" for tones between 150 and 300 d.v./sec. They then subjected the ear to a continuous tone for about two minutes, which length of time they found gave the maximum effect for the low intensities of sound used, and again measured the "persistence." They used two methods: (1) fatiguing with a constant tone and measuring the effect on tones varying over a range of about ± 15 d.v./sec. from the fatiguing tone; (2) varying the fatiguing tone over a similar range and measuring its effect on the persistence of a constant tone.

The two methods gave similar results and showed an increase of persistence, which was a maximum for the tone used, and decreased on either side to zero when the difference of frequency exceeded ±8 d.v./sec. approximately. The effect was greater the higher the frequency, both in the range of frequencies affected and in the amount by which the persistence was increased. Persistence was found to increase with the intensity of the tone; but with these higher intensities there was a decrease in the fatigue effect, i.e., the "increase of persistence" as a result of fatigue tended to disappear. It is suggested that this is because the louder intensities produce fatigue in a much shorter time, even during the time in which the measurements of the critical frequency of pulsation were being made. Hence no further effect is to be expected.

One is impressed by the various results obtained. There is some loss of ability to respond when the ear is subject to a continuous stimulation, but there is no agreement as to whether this is confined to the pitch of the continuous sound or whether it is general and affects all frequencies equally.

It is to be noted that when the effect is found to be specific the same ear has been used for comparison before and after stimulation; when a general factor is found comparison has been made with the previously unstimulated ear. This may be one cause of the discrepancy, the specific effect being due to some change in the cochlea, the general to a muscular factor. If so, the muscular factor should disappear if both ears are stimulated. This latter question has been partially tested by Bartlett and Mark (7), who found that stimulation of both ears with the same continuous sound, but with a certain time difference which gave localization to one side, produced a set to one side, as reported by Flügel. (Bartlett and Mark did not test whether the effect was general.) This finding, which has been confirmed by Pattie (38), would seem to shift the locus from the ear to the higher centers, so that the claims of those who maintain that adaptation to noise is a central factor are not without the confines of probability. If so, adaptation way well be an active process making a certain demand on the organism and producing eventually a state of general and widespread fatigue. This would be evidenced by a great sense of strain and a craving for quietude, such as is shown by many town dwellers and by others when the general bodily tone is low. In these circumstances, it is not that the sensory processes are more acute; they may be less acute; it is the ability to inhibit the passage of the sensations into consciousness which fails, causing a further drain on the organism which may snowball on till, eventually, there is a collapse.

REFERENCES

1. Achmatov, A. S. Ueber die aukustische Adaption. *J. f. Psychol. u. Neur.*, 1925, **31**, 428-442.
2. Banister, H. Auditory theory: a criticism of Professor Boring's hypothesis. *Amer. J. Psychol.*, 1927, **38**, 436-440.
3. ———. Phase effect and the localisation of sound. *Phil. Mag.*, 1926, **2**, 402-431.

4. ————. A suggestion towards a new hypothesis regarding the localization of sound. *Brit. J. Psychol.*, 1926, **17**, 142-153.

5. ————. Three experiments on the localization of tones. *Brit. J. Psychol.*, 1926, **16**, 265-292.

6. ————. The transmission of sound through the head. *Phil. Mag.*, 1926, **2**, 144-161.

7. Bartlett, F. C., & Mark, H. A note on local fatigue in the auditory system. *Brit. J. Psychol.*, 1922, **13**, 215-218.

8. Bishop, H. C. An experimental investigation of the positive after-image in audition. *Amer. J. Psychol.*, 1921, **32**, 305-325.

9. Bode, R. Die Zeitschwellen für Stimmgabeltöne mittlerer und leiser Intensität. *Psychol. Stud.*, 1907, **2**, 293-323.

10. Bunch, C. C. Measurement of auditory acuity throughout the tonal range. *Psychol. Monog.*, 1922, **31**, 45-82.

11. Crandell, I. B. Sounds of speech. *Bell Sys. Tech. J.*, 1925, **4**, 586-626.

12. Eberhardt, M. Ueber die phälnomenale Höhe und Stärke von Teiltönen. *Psychol. Forsch.*, 1922, **2**, 346-367.

13. Fletcher, H. The nature of speech and its interpretation. *J. Franklin Instit.*, 1922, **193**, 729-748.

14. ————. Physical measurements of audition and their bearing on the theory of hearing. *J. Franklin Instit.*, 1923, **196**, 289-326.

15. ————. *Bell Sys. Tech. J.*, 1923, **2**, 165.

16. ————. The physical criterion for determining the pitch of a musical tone. *Phys. Rev.*, 1924, **23**, 427-437.

17. Fletcher, H., & Wegel, R. L. The frequency sensitivity of normal ears. *Phys. Rev.*, 1922, **19**, 553-565.

18. Flügel, J. C. On local fatigue in the auditory system. *Brit. J. Psychol.*, 1920, **11**, 105-134.

19. Fowler, & Wegel, R. L. *Phys. Rev.*, 1922, **19**.

20. ————. *Trans. Amer. Laryng., Rhino., & Otol. Soc.*, 1922.

21. Hancock, C. The effect of the intensity of sound upon the pitch of low tones. *Psychol. Monog.*, 1914, **16**, 161-165.

22. Helmholtz, H. L. F. On the sensations of tone. (Trans. by A. J. Ellis.) (3rd ed.) London: Longmans, 1895. Pp. xix+576.

23. Kingsbury, B. A. A direct comparison of the loudness of pure tones. *Phys. Rev.*, 1927, **29**, 588-600.

24. Knudsen, V. O. Sensibility of the ear to small differences of intensity and frequency. *Phys. Rev.*, 1923, **21**, 84-102.

25. Knudsen, V. O., & Jones, I. H. *Ann. Otol., Rhinol., & Laryng.*, 1925.

26. ————. *Laryngoscope*, 1926.

27. Kucharski, P. Sur la persistance des sensations auditives. *C. r. Soc. biol.*, 1927, **97**, 691-693.

28. Lane, C. E. Minimum sound energy for audition for tones of high frequency. *Phys. Rev.*, 1922, **19**, 429-497.

29. Luft, E. Ueber die Unterschiedsempfindlichkeit für Tonhöhen. *Phil. Stud.*, 1888, **4**, 511-540.

30. MacKenzie, D. The relative sensitivity of the ear at different levels of loudness. *Phys. Rev.*, 1922, **20**, 331-348.

31. Mader, L. Mikrophronische Studien am schallleitenden Apparate des menschlichen Gehörorganes. *Sitzber. Akad. Wiss., Wein, Math.-naturw. Kl* (III. Abt.), 1900, **109**, 37-38.

32. Mayer, A. M. Researches in acoustics. *Amer. J. Sci.*, 1894, **47**, 1-28.

33. Meyer, M. Ueber die Unterschiedsempfindlichkeit für Tonhöhen nebst ein-

35. Myers, S. C. A text-book of experimental psychology. (2nd ed.) London: Cambridge Univ. Press, 1911. Vol. I, pp. 344; Vol. II, pp. 107.

36. Nesfield, —. Deafness and its alleviation. 1928.

37. Nikiforowsky, P. M. Der Abfluss der akustischen Energie aus dem Kopfe, wenn ein Schall durch die Stimme oder durch den Diapason-Vortex zugeleitet wird. *Zsch. f. Sinnesphysiol.*, 1912, **46**, 179-197.

38. Pattie, F. A., Jr. An experimental study of fatigue in the auditory mechanism. *Amer. J. Psychol.*, 1927, **38**, 39-58.

39. ————. *Brit. J. Psychol.*, 1928, **19**.

40. Pohlman, A. G., & Kranz, F. W. Observations on hearing-acuity for bone-transmitted sound. *Proc. Soc. Exper. Biol. & Med.*, 1927, **24**, 456-461.

41. Rayleigh, (Lord). Acoustical notes. Sensations of right and left from a revolving magnet and telephones. *Phil. Mag.*, 1907, **13**, 316-319.

42. Révész, G. Ueber musikalische Begabung. *Ber. d. VI. Kong. f. exper. Psychol.*, 1914, 88-90.

43. Riesz, R. R. Differential intensity sensitivity of the ear for pure tones. *Phys. Rev.*, 1928, **31**, 867-875.

44. Sacher, A. Beitrag zur Lehre der professionellen Schwerhörigkeit. Die Taubheit der Kesselschmeide. *Monatssch. f. Ohrenhk.*, 1927, **61**, 337-359.

45. Schäfer, —. *Passow u. Schäfer'sche Beitr.*, 1913, **6**, 207.

46. Sewall, E. Beitrag zur Lehre von der Ermüdung des Gehörorgans. *Zsch. f. Sinnesphysiol.*, 1907, **42**, 115-123.

47. Smith, E. M., & Bartlett, E. C.. On listening to sounds of weak intensity. *Brit. J. Psychol.*, 1917, **10**, 133-168.

48. Stewart, G. W. The function of intensity and phase in binaural location of pure tones. *Phys. Rev.*, 1920, **15**, 425-455.

49. Stucker, N. Ueber die Unterschiedsempfindlichkeit für Tonhöhen in verschiedenen Tonregionen. *Sitzber. Akad. Wiss., Wien, Math.-naturw. Kl.*, 1907, **116**, 367-368.

50. Stumpf, C. Tonpsychologie. Leipzig: Hirtzel, 1890. Vol. II. Pp. xiv+582.

51. Travis, L. E. Changes in auditory acuity during the performance of certain mental tests. *Amer. J. Psychol.*, 1926, **37**, 139-142.

52. Urbantschitsch, V. Ueber das Anklingen acustischer Empfindungen. *Pflüg. Arch. f. d. ges. Physiol.*, 1881, **25**, 323-342.

53. Valentine, W. L. Note on the "binaural beat." *J. Comp. Psychol.*, 1927, **7**, 357-368.

54. Vance, T. F. The lower limit of tonality. *Psychol. Monog.*, 1914, **16**, 115-149.

55. Watt, H. J. The psychology of sound. London: Cambridge Univ. Press, 1917. Pp. vii+241.

56. Wegel, R. L. The physical examination of hearing and binaural aids for the deaf. *Proc. Nat. Acad. Sci.*, 1922, **8**, 155-161.

57. Weinberg, M., & Allen, F. On the critical frequency of pulsation tones. *Phil. Mag.*, 1924, **47**, 50-62.

58. Wilson, H. A., & Myers, C. S. The influence of binaural phase differences on the localisation of sounds. *Brit. J. Psychol.*, 1908, **2**, 363-385.

59. Young, P. T. Auditory localization with acoustical transposition of the ears. *J. Exper. Psychol.*, 1928, **11**, 399-429.

igen Bemerkungen über die Methode der Minimaländerungen. *Zsch. f. Psychol.*, 1898, **16**, 352-372.

34. Miller, D. C. The science of musical sounds. New York: Macmillan, 1916. Pp. vii+286.

CHAPTER 7

HEARING II*

H. Hartridge
St. Bartholomew's Medical College, London

AND

H. Banister
Cambridge University

I. Anatomy of the Ear

The Physiology of Hearing. Hearing is classed with vision and smell among the distance receptor mechanisms because it conveys to our higher centers information of events taking place at a distance. It lacks the fine detail-perceiving qualities of vision, but it has the great advantage of not requiring an uninterrupted view. This is due to the fact that sound waves do not as a general rule form shadows. In consequence we hear nearly as well by the ear that is pointing away from a sound as by that which is pointing at it. If sound waves and light waves behaved in an identical manner, a speaker's vocal cords would have to be in a straight line with his throat, and he would have to point it at us. We, in turn, should have to aim our ear at him, or we should not hear him. In this and other important ways, sound and light and, consequently, the eye and the ear differ. We still give vision the first place as a distance receptor, but the ear is more than keeping pace with its rival. The eye has acquired the telescope and microscope, it uses photography and X-rays. The ear, in no way outclassed, has acquired the gramophone and wireless. By our ears, we can safely steer our ships into harbor in the densest fog or can locate hidden rocks or submarines. By our ears, we can learn the position of aircraft up in the clouds beyond the range of vision. In ways such as these, the ear is the keen competitor of the eye. And thus it becomes its most successful ally.

While practically all observers are agreed on the fundamental aspects of hearing, there are marked differences of opinion about the mechanical and physiological details. It is no reflection on the ability, enthusiasm, and enterprise of workers in this field that this is the case. The matter for surprise is that we know as much as we do. The difficulties in the way of the experimenter are well-nigh insuperable. The organ of hearing is placed 3 to 4 cms. deep to the surface of our skulls. Above it lies the cranial cavity with its delicate contents. Below it are found the structures of pharynx and jaw. Its more delicate parts are situated in some of the hardest bone in the body, and the whole mechanism is extremely small. Histological examination is consequently difficult. Observation on the different parts at work is quite impossible. This explains why rivalry exists between theories. It is the nature of man's mind that he must needs try to picture with his imagination what he cannot see with his eyes.

---

*Sections I-V and VIII by Hartridge; Sections VI and VII by Banister.

The pinna is an irregular plate of yellow elastic cartilage covered with skin. Muscles and fibrous bands attach it to the head. Marked individual variations in shape occur. In some lower animals it probably plays an important part as a collector of sound waves. It can be erected or "pricked." It can be turned in different directions so as to assist in determining the locality of the source of sound. It has an apparatus at its base for varying the size of the aperture through which sounds can pass. These protect the internal mechanism from possible damage caused by loud local sounds. In man, the pinna is in a state of retrogression. Few of us can move our ears more than a few millimeters. None of us can "close" our ears, except metaphorically.

The effect of the pinna on hearing is problematical. While its removal is said by some (39) to produce no effects, McKendrick (35, Vol. II, p. 672) found that covering it with plaster of Paris weakened intensity and altered quality. It has been suggested that it acts as a resonator (25) and as a reflector (83, p. 416) for tones of high pitch.

The external acoustic meatus is a bent tube about 3 cms. long. Its first or cartilaginous part slopes downwards. Its second or osseous part slopes downwards and forwards and inwards. It is lined with thin, closely adherent skin. Stiff hairs grow from this and point towards the entrance. Ceruminous glands secrete a bitter wax. The hairs, the wax, and the shape and narrowness of the meatus discourage the entry of foreign bodies. The effects of the external acoustic meatus on hearing is to act as a conductor to sounds of all pitches. Politzer (29) thought that it reflected the sound waves on to the drum-skin, but this is extremely doubtful.

The tympanic membrane or drum-skin is a pearly gray structure nearly 1 cm. in diameter, made of radial and circular fibers of connective tissue. On the outer side it is covered with very thin skin. Its inner surface is covered with mucous membrane. Its shape is markedly conical. It is mounted at the end of the external acoustic meatus at a very oblique angle. To its upper quadrant is attached the handle of the malleus. This may be seen through the meatus as a reddish streak.

The function of the drum is to convert into movements the variations of pressure of the air outside the head which are conveyed down the external acoustic meatus. This it does without giving preferential treatment to any particular part of the audible range of tones because of its conical shape.

It should be noted, however, that this conical shape has the disadvantage of producing summation and difference tones when two or more tones are simultaneously presented to the ear. Thus, if pure tones of 100 and 150 d.v./sec. fall on the drum, the vibrations performed by the ossicles would correspond to these two tones, and there would be, in addition, a difference tone of 50 d.v./sec. and a summation tone of 250 d.v./sec. (5, p. 391).

The middle ear is an air-containing cavity surrounded by bone on all sides, the bone being lined by periosteum and mucous membrane. To the walls of the middle ear are attached the chain of ossicles which convey the sound vibrations from the drum-skin to the internal ear. The middle ear communicates with the nasal cavity by means of the Eustachian canal. The nasal end of this is closed by means of a valve, which is temporarily opened during the act of swallowing. At these times the pressure of the air inside the middle ear is allowed to become equal to that in the nasal cavity, which is practically identical with that outside the head. The pressure of air on either side of the drum is thus permitted to become the same. Divers and workers in caissons have to remember to swallow periodically while the pressure of the air is being increased or decreased. It is stated that sudden decompression with a closed Eustachian canal may cause permanent damage to the internal ear. The names of the three ossicles are the malleus (or hammer), the incus (or anvil), which articulates with the malleus, and the stapes (or stirrup-bone), which articulates with the incus and conveys the vibrations by its footplate to the oval window of the internal ear. The ossicles are clothed by a periosteum which is covered with mucous membrane. The way that vibrations imparted to malleus are conveyed through incus to stapes may be very briefly outlined as follows. Between malleus and incus is a saddle-shaped articulation, and on malleus is a spur of bone which engages with the body of incus. In consequence, when the bones are in their normal positions, they vibrate practically as a single compound lever which is free to rotate about a fulcrum passing through the lateral processes of the two bones. Movements of the drum applied to the handle of malleus will thus cause corresponding but smaller movements of the long process of incus. The lower end of this carries a ball-and-cup joint, the head of stapes forming the ball. The vibrations are thus imparted to the footplate of stapes which fits the oval window in a watertight manner by means of its annular seal. In addition to the three ossicles, the tensor tympani and stapedius muscles must be mentioned. The former is innervated by a branch of the trigeminal or Vth cranial nerve. When stimulated it tends to increase the tension in the fibers of the drum-skin. Contraction normally occurs in the tensor muscles of both ears if musical tones or noises fall on either ear. Those of high pitch are said to be particularly efficacious. The stapedius muscle is innervated by a branch of the facial, or VIIth cranial nerve. Some observers hold that it antagonizes the effect of the tensor tympani, others that it reduces the pressure on the contents of the internal ear; Hartridge's view is that it prevents chattering between the ossicles and thus avoids the loss of weak vibrations. The movements of the ossicles have been observed in a number of ways. Politzer (30) attached glass threads to the ossicles and recorded the movements on a smoked drum. Burnett (7, p. 95) opened the middle ear, sprinkled powdered starch on the ossicles, and examined their movements when the drum was set into vibration by means of organ pipes. He records that the movements were quite small but easily observable. In addition, he noted vibrations in the membrane covering the round window (which will be described in detail very shortly) and saw that these were nearly equal to the corresponding movements of the stapes. Helmholtz himself (19) observed the movements of the ossicles and found that they were less than a tenth of a millimeter. Frey (13) states that he has not been able to observe them. It is possible that the tones he employed for setting the drum into vibration were not of sufficient intensity to produce movements large enough for visual observation. The natural period of vibration of the mechanism represented by drum, ossicles, and muscles has been investigated by Everett (9, p. 94), who states that it is about 10 d.v./sec. This is well below the lowest audible musical tone, which is approximately 16 d. v./sec. Consequently, some tones in the audible range are not favored by the mechanism at the expense of other tones. The area of the footplate of the stapes is roughly one-twentieth that of the drum-skin. There is probably a gain in pressure with a corresponding loss of volume as sounds are transmitted from the external auditory meatus into the internal ear. This has been variously estimated at from 10 to 30 times. The lower figure is probably the more likely, particularly for tones of high pitch. It should be noted that in birds the chain of ossicles is replaced by the single ossicle, the columella, which is attached at one end to the drum-skin and at the other end projects into the internal ear, thus communicating the sound vibrations to the latter.

The internal ear lies in a cavity hollowed out from solid bone. This cavity contains a number of membranous structures, parts of which belong to the mechanism for equilibration. The part which concerns hearing is called the cochlea. It consists of a conical mass of bone, the modiolus. about which is wound a spiral tube, the cochlear canal. This tube is divided by two membranes, the basilar membrane and Reissner's membrane, into three canals. The upper or scala vestibuli is in communication with the fluid which bathes the end of the footplate of the stapes. The middle canal, or scala media, is filled with a second fluid, and communicates with the membranous structures which fill the rest of the internal ear and are used for equilibration. The lowest canal or scala tympani communicates at its upper end with the upper canal through a minute opening, the helicotrema. Its lower end terminates blindly, just opposite the round window. This would communicate with the middle ear if it were not closed by a membrane. Reissner's membrane is very thin and plays no important part in the perception of tones other than effecting a separation between the two liquids in the upper and middle canal. The lower or basilar membrane carries structures of great importance to hearing. Its inner edge is attached very firmly to a spur of bone, the bony spiral lamina which projects from the modiolus. Its outer edge is attached to a fibrous and vascular structure called the stria vascularis. Its upper surface carries a series of arches, the arches of Corti; on either side of these are mounted supporting cells, which have lying in between them long sensory cells, which carry at their upper surfaces a number of fine hairs. For this reason, these cells are called "hair-cells." Nerve-fibers which are connected with these hair-cells run through ganglia in fine channels hollowed out in the modiolus. These unite to form the auditory portion of the VIIIth cranial nerve. On the lower side of the basilar membrane are supporting cells only.

When the piston of the stapes makes in-and-out movements under the influence of vibrations set up in the drum-skin by sounds, corresponding up-and-down movements of the basilar membrane occur with corresponding up-and-down movements of the hair-cells mounted on either side of Corti's arches. Since the whole internal ear is filled with liquid and since all fluids are incompressible, the up-and-down movements of the basilar membrane must be accompanied by the corresponding movements of some structure in communication with the lower scala tympani. This structure is the membrane closing the round window. When the piston of the stapes vibrates and thus causes the basilar membrane to vibrate, corresponding vibratory movements are set up in this membrane, as Burnett (7, p. 95) observed.

We must now consider how up-and-down movements of the hair-cells causes impulses to travel from them up the fibers of the auditory nerve. Above the basilar membrane is mounted the tectorial membrane. The inner edge of this is attached to the bony spiral lamina, while its outer edge is free. The body of this membrane lies horizontally immediately above the tips of the hair-cells. Helmholtz thought that the tips of the hairs touched the under surface of the membrana tectoria whenever the basilar membrane moved upwards. Wrightson (47) thought that the tips of the hairs were embedded in the lower surface of the tectorial membrane and that the hairs were bent first in one direction and then in the other as the basilar membrane rose and fell. Hartridge's view is that during the ascent of the basilar membrane the upper surface of the supporting cells presses against the lower surfaces of the tectorial membrane, causing the two to rise together. When, however, the basilar membrane comes to rest and then commences to fall, the tectorial membrane is caused to undergo precisely similar movements by the pulling of the hairs which are attached to its under surface. Whatever may be the details, most physiologists agree that movements of the basilar membrane bring about stimulation of the hair-cells so that impulses are caused to travel up the fibers of the auditory nerve. Hardesty (16, p. 109) believes that the tectorial membrane is set into vibration by the movements of the piston of the stapes and that in consequence stimulation of the hair-cells occurs. It is not surprising that these differences of opinion exist, because no one has ever witnessed the mechanism in operation; for, as I have indicated in the introduction, the difficulties of making such observations are insuperable.

II. Theories of Hearing

Numerous hypotheses have been advanced to explain how hearing results from the vibrations set up in the basilar membrane and the corresponding stimulations of the hair-cells. We may effect an initial subdivision of these hypotheses under the following headings:

1) *Telephone Theories.* These theories hold that all parts of the basilar membrane vibrate with approximately equal amplitude at any one moment, the amplitude being determined by that of the piston of stapes, which in its turn depends on the movements of the drum-skin. According to this view, the basilar membrane and its hair-cells perform the same kind of function that a microphone does. The analysis of the impulses which travel up the auditory nerve, according to the frequency and intensity of the sounds falling on the ear, has to be performed by the central nervous system. This is called the "telephone hypothesis." It has been associated very largely with the names of Rutherford (34) and Wrightson (47). Against this hypothesis we have the difficulty that the highest audible musical note is probably considerably above 10,000 d.v./sec. In consequence, if a note of this frequency is to be differentiated from other notes, the vibrations of the basilar membrane must occur at this rate and must cause corresponding impulses to travel up the auditory nerve to the central nervous system. Experimental evidence seems to point to 600 or 700 impulses per second as the maximum number of impulses per second that a mammalian sensory nerve is capable of conducting. For this fundamental reason, this theory has had but little support of recent years.

2) *Sound-Pattern Theories.* When metal plates are suitably clamped and are set into vibration by striking or bowing, the observations on the patterns made by sand which has been scattered over the surface indicate the mode of vibration of the plate. This beautiful phenomenon has been made the basis of a theory of sound analysis by the cochlea. It is supposed that sounds falling on the ear set up vibrations, which are communicated by the ossicles to the internal ear and set selected parts of the basilar membrane into vibration, other intermediate parts being stationary. Further, according to the theory, as the pitch and intensity of the incoming sounds alter, so also do changes occur in the modes of vibration of the basilar membrane. In consequence, the impulses travelling up the auditory nerve to the central nervous system vary according to the sounds which are falling on the ears. The recognition of these groups of impulses permits the higher centers to identify the sounds which are being received. This is called the "sound-pattern theory." It has been associated very largely with the names of Waller (49) and Ewald (10). The latter has made very ingenious models which permit vibrations of a rubber membrane of small dimensions to be observed under the microscope. The mode of vibration was observed to change when changes were made in the sounds which were causing the membrane to vibrate. There are no fundamental difficulties against this hypothesis. The main objection to it may be explained as follows. According to Ewald's model the vibrations of the membrane for a single pure tone were of quite a complicated nature. If,

then, a number of tones fell on the ear at the same time, the sound patterns produced in consequence would be of an extremely complicated nature, for not only have the fundamental tones to be perceived but also the partial tones (overtones) which accompany each of the fundamentals. We shall see later that these partials vary in intensity according to the musical instrument which is producing them. The different vowel sounds are produced, moreover, by the human voice by variously accentuating these different partials. If the ear is to recognize different musical instruments or different vowels, it must correctly estimate the intensity of these partial tones. Therefore, these must contribute their share to the sound pattern in addition to the fundamental tones. The sound pattern will in consequence become of so complicated a nature that it is difficult to see how the mechanism could effect the detailed analysis which experiment and everyday experience demonstrates to be one of its outstanding properties.

3) *Sound-Wave Theories.* If by suitable means, waves are caused to travel along a membrane, it is found that the distance between the crests decreases as the note increases in frequency. This fact has been made the basis of theories of hearing. One of these supposed that movements of the piston of stapes caused waves to travel up Reissner's membrane. When they reached the top of the cochlea, they returned along the basilar membrane, the distances between the crests being a measure of the frequency of the tone setting the drum-skin into vibration. This theory was advanced by Hurst (20). A somewhat similar theory was advanced by Max Meyer (27). He supposed the vibrations to travel up the basilar membrane, but the waves suffered a decrement which tended to make them die out as they travelled towards the upper end of the cochlea. In consequence, the waves produced by loud tones might almost reach the top, whereas those produced by tones of low intensity would travel but a short distance. In this way, the loudness of a tone would depend on the number of hair-cells affected by it during its passage. The greater the intensity of the tone, the greater the number stimulated. Boring (6, p. 157) has recently revived interest in theories of this nature. He particularly says that he is not proposing a theory of hearing; he does, however, critically examine all other possible theories and excludes them one at a time as being unlikely. It would appear justifiable, therefore, to connect his name with theories which depend primarily on intensity analysis. The outstanding difficulty in the way of their acceptance is that they require tones of all pitches to be transmitted without frequency alteration up the auditory nerve. They, therefore, have the same fundamental objection against them that confronts the simpler telephone theories of Rutherford and Wrightson from which they have developed.

4) *Volume Hypothesis.* As one travels down the musical scale, one observes the change in pitch, but according to Watt there is in addition what he calls an increase in volume. According to his view, a low-pitched tone has a large volume, whereas a high-pitched one has only a small one. "Low tones are great and massive. High tones are sharp and thin." He pictures low tones as acting as stimuli for large areas of basilar membrane.

In consequence, large numbers of nerves receive impulses from their hair-cells. On the other hand, few nerve-fibers receive impulses in consequence of high-pitched tones. Further details of this hypothesis and criticism of it will be found in Section VI.

5) *Resonance Hypotheses.* If one presses the loud pedal of a piano so that the dampers are lifted from the wires, and some musical instrument which will play a sustained tone be sounded, on terminating the tone the piano wires will be heard to continue to sound it. Further experiment demonstrates the fact that certain wires have been set into sympathetic vibration or resonance by the energy they receive from the source of sound and that a portion of this energy which they have stored is returned to the air again as a continuation of the same musical tone. This simple experiment illustrates the fundamental principle on which the resonance theory of hearing is based. Briefly, it regards the basilar membrane together with the structures mounted on it as forming a series of resonators which have natural frequencies of vibration corresponding roughly to all the musical tones. When a particular tone falls on the ear and is conveyed by the ossicles to the fluid of the internal ear, the hypothesis supposes that a localized area of the basilar membrane is set into vibration, and thus stimulation of the corresponding hair-cells causes impulses to travel up those fibers of the auditory nerve which belong to that localized area. A tone of higher or lower pitch will, according to this theory, set a different area of basilar membrane into vibration and will thus cause impulses to travel up a different set of auditory nerve-fibers. If the tones are accompanied by partials, then in addition to the area of basilar membrane corresponding to the fundamental tone, other areas of basilar membrane which correspond to each of the partial tones will be set into vibration. In just the same way that we can strike certain selected keys of the piano with our fingers and set the corresponding strings into vibration, so vibrations, when they fall on the ear, will, according to the theory, set selected areas of basilar membrane into vibration, the result being that impulses travel up the auditory nerve-fibers corresponding to those areas and reach the auditory area of the brain. Variations in intensity are perceived partly by variations in the width of the area of basilar membrane affected, and therefore by variations in the number of auditory nerve-fibers which convey impulses to the central nervous system, and partly by variations in the number of fibers which receive stimulation from any one localized part of the basilar membrane. In consequence, a just-audible tone may stimulate an extremely small number, possibly only one fiber of the auditory nerve with impulses of an all-or-none nature which have considerable intervals between them, while a very loud sound indeed may cause impulses from an enormous number of auditory nerve-fibers to pass in continuous streams to the auditory area. Detailed consideration on these and other points will be given in Section VIII of this chapter.

III. Physiological Evidence

The experimental evidence is found to fall under the following headings: anatomical and histological evidence; the results of pathological investigations; experiments on animals; the results given by mechanical models; evidence of a physical nature. Each of these different categories will be considered in the sections which follow.

Anatomical and Histological Evidence. If a series of sections of the cochlea be examined, it will be seen that marked differences occur as one passes up the cochlear canal. The canal itself is definitely larger at the bottom than it is at the top. The bony spiral lamina also is notably wider at the bottom than it is at the top, so also in the mass of tissue attached to the opposite wall of the cochlear canal, namely, the stria vascularis. But the increase in width as one passes down the cochlear canal affects the bony spiral lamina to a far greater extent than it does the canal itself. In consequence, the basilar membrane which is stretched across from the tip of the bony spiral lamina to the stria vascularis decreases in width as one passes from the top of the canal to the bottom. That is, it alters its dimensions in precisely the opposite direction to those of the other three structures we have been considering. Thus, the shortest fibers of the basilar membrane are those nearest to the oval and round windows. If the structures attached to different parts of the basilar membrane be carefully examined, it will be seen that they vary greatly in mass at different parts. The shortest fibers of the basilar membrane have small cell masses attached to them. On the other hand, the long fibers of the basilar membrane are found to be very heavily loaded with cells. If we examine the sectional areas of these cell masses attached to the shortest and longest fibers in a number of sections, we shall find that they suggest that the dimensions increase more than do the lengths of the fibers to which they are attached. Consequently, if we assume, for the sake of argument, that the short fibers are one-third the length of the long ones, then the cell masses should be in the ratio of 1 to 9, if their dimensions corresponded precisely to those of the basilar membrane fibers to which they are attached. But instead they appear to vary by a distinctly greater amount than this, possibly by as much as 1 to 12. On scrutinizing the bony spiral lamina and the stria vascularis, we find corresponding changes in their massiveness. Those near the bottom of the cochlear canal which are supporting the short, lightly loaded fibers are wide and strong and give the appearance of considerable strength. On the contrary, if the corresponding parts of the upper canals be examined, the bony spiral lamina and stria vascularis which have to support the long, heavily loaded fibers are seen to be relatively weak. These observations appear at first sight to be contradictory to one another. Long, heavy fibers which would be expected to require strong supports apparently do not have them, while short, lightly loaded fibers have supports provided out of all proportion to them. According to the resonance theory, the clue to this apparent contradiction lies in the degree of tension which exists in the different parts of the basilar membrane. Thus, it is necessary for the short fibers to respond to notes of very high pitch; that is, in addition to being short and lightly loaded they must have a very considerable tension. On the other hand, the long, heavily loaded fibers have, according to the theory, the function of responding to the lowest tones of the musical range. They will consequently have very low tensions and thus will require very weak lateral supports. These histological differences have their most rational explanation on the basis of the resonance theory. There would, however, be little difficulty in so modifying the alternative theories that they fitted in with some, if not most, of the histological facts.

Evidence of a Pathological Nature. Many references are made in the textbooks to the disease which boiler-makers acquire. It is shown by deafness to high-pitched tones, with some loss of acuity for tones of all parts of the audible range. An examination of the cochleae of sufferers from this disease has demonstrated in most cases lesions in the basilar membrane or the structures attached to it, which have been limited to the short fibers situated at the lower part of the cochlear canal. Several of the theories which have been outlined above can be made to account adequately for this disease. It fits the resonance theory, since high-pitched tones to which the patient is deaf are due to resonant vibrations in the short fibers, which are found to be pathologically affected. It can also be made to fit in with the telephone theory, particularly its more recent modifications, on the ground that the parts diseased are those nearest to the piston of the stapes and therefore the first to be subjected to the full force of the intense noises of the boiler shop. The sound-pattern theory of Ewald is not in this favorable position. It seems difficult to understand why the diseased areas should be limited to any particular part of the cochlear canal.

The existence in individuals of lacunae or islands of deafness has been frequently reported. In order that these islands may be disclosed, careful testing of each ear by itself has to be performed because the law of chance makes it unlikely that pathological changes affecting localized portions of the cochleae of the two ears would affect identical parts. The findings are that a particular part of the audible range of tones is missing in one ear (12). It is stated that this has in certain cases been found associated with local pathological changes in the cochlea of approximately the region to be expected on the resonance hypothesis. Such a correlation fits in admirably with the theory referred to and is, so far as can be judged, equally inexplicable on any of the other theories. Unfortunately, precise references to the papers reporting these cases have not proved traceable, although references to them are frequently made (6, p. 183).

Another pathological change frequently referred to in the literature is known as double disharmonic hearing (diplacusis). In this condition, the two ears respond to a single musical tone as if two tones out of tune with one another were being received, one by one ear, the other by the other. Thus, on stopping the external meatus of one ear, the phenomenon largely disappears. It can be easily accounted for on the resonance theory by supposing that a part of the cochlear canal of one ear has suffered such a pathological change that the basilar membrane or the structures applied to it (or both) vibrate at a different frequency from the normal, either by changes in tension, mass, or length. The result would be that a pure tone would set into vibration a different area of basilar membrane from that which it ordinarily did. In consequence, the auditory nerve-fibers conveying impulses from one ear would not correspond to those from the other and thus disharmonic hearing would result. The phenomenon would not be difficult to explain on Ewald's sound-pattern theory, since pathological change to one cochlea might cause the sound patterns to be modified in design. It is, however, difficult to reconcile with the telephone or wave theories of hearing (23, 36).

Lastly, it is stated that hearing has persisted in spite of complete pathological destruction of both cochleae. If this statement were authenticated, it would point to some other structures, in addition to or instead of the cochlea, being responsible for the recognition of sounds. This would seem to be extremely unlikely.

The Evidence of Animal Experiments. Two different kinds of experiment have been performed on animals which are related to the pathological changes referred to in the previous section. In the first variety of experiment, localized destruction by surgical interference of a limited part of both cochleae has been attempted. On the subsequent recovery of the animal, experiments to determine the existence of deafness to certain parts of the audible range have been carried out. Thus, Baginsky (1, 2, 3) is reported in Schäfer's textbook (35, p. 1181) to have found deafness to low tones after destruction of the upper parts of the cochleae of dogs while destruction of the basal portions produced deafness to high tones. Stepanoff (37) is reported by the same author to have found no loss of perception for low tones when in guinea pigs the upper parts of the cochleae were destroyed. He "tested hearing by the doubtful expedient of observing a reflex movement of the auricle when a sound wave was heard, and he found that the animals reacted even after the perilymph and endolymph were flowing out of the broken cochlea apex. This is too rough an experiment and does not appear to have much value." It should be noted that great care must be used in experiments such as these to employ sources of sound which are unaccompanied by upper partials (overtones) because, unless this is done, the partials of a low tone may be perceived and produce a reaction while the conditions are such that the tone itself cannot receive recognition. The second variety of experimental work on animals depends on the production of analogous conditions to those met with in boiler-makers, that is, the animals are subjected first of all to long-continued musical tones in the hope that pathological changes will be set up in their cochleae. This is subsequently sought for by histological examination. The most frequently referred to experiments of this nature were performed by Yoshii (48). He exposed guinea pigs to continuous tones for intervals of the order of 50 to 60 days. "He used an Edelmann whistle of 4138 vibrations, a siren of 1034 vibrations, a whistle of 976 vibrations, and an organ pipe of 193 vibrations. After exposure to one of these stimuli, the animal was killed and the cochleae examined histologically. In general, he found destruction of the organ of Corti over a range of one-quarter to one-half a turn of the cochlea. Moreover, he found that the regions of destruction occurred in the order of frequency, with the highest frequency towards the base of the cochlea." The above quotation is taken from Boring's paper (6, p. 184). The pitches of the tones are in each case followed by a mark of interrogation. The author does not explain why. To continue the reference: "Since the small end of the basilar membrane lies at the base of the cochlea, the results might seem to support Helmholtz' resonance theory, although Yoshii himself concludes that the region affected is too broad for Helmholtz' theory and falls back upon Ewald's." Boring now follows this quotation with an explanation according to his own theory of hearing, although he has said four pages previously: "I am not, let me say again, proposing a theory." With regard to the wide range of cochlea affected in Yoshii's experiments, one of the commonest misunderstandings of Helmholtz' theory is to think that a small area of basilar membrane is alone affected by any one pure tone. It would seem that the size of the area is a function of intensity and that a loud tone may affect quite a wide area of basilar membrane. It should be noticed that Yoshii used whistles, an organ pipe, and a siren. With regard to the two former, it is well known that their pitches vary with change of temperature and barometric pressure. It is also known that the fundamental tone is usually accompanied by one or more moderately strong overtones. The siren also is usually a source of variable pitch which is accompanied by overtones. It would appear justifiable to conclude that to some extent, at all events, the width of basilar membrane affected was due to variations in the pitches of the stimulating tones. When these points have been taken into account, Yoshii's experiments would appear to be more readily correlated with the resonance hypothesis than with any one of its rivals. Thus, a sound-pattern theory would necessitate a number of localized areas of degeneration. These were not found. The telephone theory would necessitate degeneration of the whole of the cochlear apparatus and not of selected localized parts. According to the wave theory, it should presumably be lower parts of the basilar membrane or the structures attached to it which suffer the pathological change. Boring has, however, accounted for their actual positions on his modification of this theory, as has been mentioned above. One of the most crucial tests so far as animals are concerned would appear to be one based on Pavlov's method of conditioned reflexes. Localized islands of deafness produced either surgically or as the result of stimulation by pure tones could readily be identified with considerable precision by this technique. It is reported that such experiments have already been performed either by Pavlov or his co-workers. No reference to them can be found, however, in Anrep's translation (31).

IV. Physical Evidence

The Evidence of Models. A number of models have been constructed in order to throw light on the theory of hearing. Helmholtz is reported to have made one. McKendrick describes in Schäfer's textbook (35, p. 1182) a model that he constructed. Models have also been constructed by Ewald (25, Vol. IV, p. 241) and Wilkinson (43, 46). With regard to Ewald, a brass box was fitted with a diagonal partition, in which had been cut a long narrow slot. This was covered by a thin rubber membrane, and the apparatus was filled with fluid. By suitably arranged glass windows, microscopic observation could be made of the vibrating parts of this membrane. Ewald obtained with it evidence of an indisputable nature pointing to a sound-pattern hypothesis of hearing. Wilkinson's models differed from Ewald's in employing a number of stretched wires to represent the fibers of the basilar membrane. Observation was made of the vibrating parts by the heaping-up of small particles of blue enamel. His papers illustrate the responses which he obtained. In his opinion, they proved two facts: (1) that the cochlea behaves as a resonator, and (2) that the mass of the fluid which fills the apparatus and is set in motion by the piston of the stapes forms part of the mass of the vibrating system of the basilar membrane. With regard to the first conclusion, four sets of photographs are in Hartridge's possession. Set 1 was reprinted from the *Journal of Laryngology and Otology* (43); Set 2 was sent privately by Dr. Wilkinson; Sets 3 and 4 were also sent privately, but were, I believe, published by the British Association in 1922. The number of photographs of the apparatus after musical tones had been applied is shown in column 2; in column 3 is shown the number definitely showing single responses, while in column 4 is shown the number definitely showing multiple responses.

TABLE 1

Set no.	No. of photographs	Definitely single	Definitely multiple
1	3	1	2
2	3	None	2
3	5	None	4
4	2	1	1

Sets 2, 3, and 4 appear to be those reproduced in Plate IV, Wilkinson and Gray's book (46, p. 85). Of thirteen tests, two alone have given definitely single responses, while nine appear to have given definitely multiple ones; two are difficult to class. The experimental evidence given by these models would appear to be more definitely in favor of the sound-pattern theory than of resonance. Up to a point, this result is satisfactory because it would be difficult to understand why one experimenter (Ewald) should obtain quite different results from another (Wilkinson). With regard to the bearing of these experiments on the mode of operation of the cochlea, I must confess to being skeptical. Both models were different in many ways from the apparatus in the cochlear canal. In one very important particular, they differ very greatly, namely, dimensions. It is a well-understood fact that scale models frequently give quite different results from their full-sized counterparts; particularly is this the case when dealing with the fluid friction aspects of hydraulics. Until a model has been constructed of precisely the same dimensions as the cochlea itself, I think it would be well to suspend judgment. With regard to Wilkinson's second conclusion, that the fluid interposed between the piston of stapes and the particular area of basilar membrane in vibration adds its load to that of the membrane, I should like to mention the following points. In Wilkinson's photographs, no single set taken of the same model shows more than one single response. One is therefore at liberty to assume that any one of the multiple responses is the true one and that all the others are false. This enables one to draw very nearly whatever conclusion one wishes. It would seem to me, therefore, to be unwise to come to any conclusion with regard to the fluid load without further evidence. Such evidence does not, unfortunately, appear to be given by a further piece of apparatus made by Dr. Wilkinson (45). My difficulty in accepting the hypothesis may be stated as follows. The piston of stapes, when it is set in motion, applies forces to the liquid in the cochlear canal. These forces are applied in all possible directions over a solid angle of 180°. The area of basilar membrane set into vibration is similarly affected by the liquid in contact with it, also through a solid angle of approximately 180°. On Wilkinson's hypothesis a more or less limited fluid path extends like a narrow curved band between the piston of stapes and the basilar membrane. I cannot see any reason for such a view.

Experiments on the Critical Frequency of Pulsation of Tone. Experiments on the critical frequency of pulsation of tone have been performed by Allen and Weinberg (42). If a rotating disc which has suitable slots cut in it be interposed between a source of sound and tubes going to the ears, it is found by experiment that a particular rate of rotation is required in order that the pulsations produced in the tones shall just disappear. If, now, prior to making each observation, the ear is stimulated for two minutes to a pure musical tone, then it is found that not only this tone but others near it in pitch are affected so that a slower rate of rotation suffices to abolish the pulsating effect. The magnitude of the effect and its spread on either side of the pure tone to which the ear is first subjected depends on the loudness and pitch and to some extent on the observer. The maximum effect is roughly 10 per cent, and it affects a range of approximately 15 d.v./sec. on either side of the pure tone. So far as is known, the critical speed of the disc measures the sensitivity of the ear to small differences of intensity. It will be shown later that these experiments give valuable evidence in favor of the resonance theory.

If we suppose that a small modification in the sound pattern alone occurs for a small change of pitch, then we can account for the phenomenon on the sound-pattern theory also. But it is very difficult to find satisfactory explanations either on the telephone theory or on the wave theory. In the case of either of these, we should expect all audible frequencies to be affected to approximately the same extent.

The Effect of Changing the Phase of a Musical Tone by a Half Wave-Length (17). If periodically applied forces are causing vibration of a mechanical system which has a natural period of vibration and if these applied forces are reversed in their time relationships so that the vibration of the mechanical system is being opposed by the forces which are now being applied to it, then the vibrations gradually die away so that the mechanical system comes to rest. Then fresh vibrations begin and are gradually built up which are exactly opposed in phase with those which the system was originally performing. For example, the case of a body of men crossing a suspension bridge and setting it swinging may be cited. If the men change step, the motion of the bridge will first be arrested and will then be restarted in time with the new step; this process can be repeated any number of times. On the other hand, if forces of sufficient magnitude are applied to a mechanical system which has no natural period so as to make it vibrate, then the change of phase in the incoming forces will be followed by an immediate reversal in the vibrations. We can use this test for ascertaining whether any mechanical system has a natural period of vibration or not. If it has a natural period of vibration, its vibrations will slowly die out and will then slowly recommence. On the other hand, if it has no natural period of vibration, its vibrations will suffer no interruption. This test can be applied to the ear with suitable apparatus (17), the principal difficulty being to devise the necessary apparatus for changing the phase of the tone which is being used for the test, without at the same time changing the intensity or quality of the tone. The following method was found to give satisfactory results. A de la Tour siren was altered so that its wind chest could be rotated about the same axis as that of the siren disc. By means of fixed stops, which were covered with rubber in order to absorb shock and so prevent the production of noise, the rotation of the wind chest was limited to that amount which would just produce a change of phase of one-half vibration in the musical tone made by the siren. Fortunately the rotation required is the same for all musical tones. In the case of the siren that was used, it was 10 degrees. A spring was fitted so as to produce the rotation, as well as a trigger mechanism for releasing it suddenly. On making observations with this apparatus, it was found that a sudden change of intensity occurred in the tone each time the trigger was released. The intensity appeared to fall rapidly to zero and then rapidly recover its original strength, or possibly it momentarily reached a somewhat greater intensity than it had had before. The phenomenon was heard equally well with all tones within the range of the siren, and it was readily heard by a large audience. Careful tests showed that the change in the intensity was not of instrumental origin. It must, therefore, have been produced by some mechanism in the ear. We are forced to conclude, therefore, that a musical tone falling on the ear sets into vibration some mechanism having a natural period of vibration, for unless this is the case no such change of intensity would have been produced. Now the drum-skin ossicles and muscles together form a mechanism which has a natural period of about 10 d.v. per second. That is, when it is transmitting a tone of, say, 250 d.v., it is performing what are called forced vibrations. It behaves to these as if it were dead beat, and therefore it is clear that this mechanism cannot be responsible for the phenomenon which has been observed. We must conclude, therefore, that the mechanism in question lies in the internal ear. Such a supposition would fit in with either the resonance hypothesis or the sound-pattern hypothesis since in both cases the assumption made is that different parts of the cochlear mechanism have different natural periods of vibration. On the contrary, the telephone hypothesis and sound-wave hypothesis make the assumption that the basilar membrane and its associated structures have no natural periods of vibration and therefore do not act selectively to different incoming tones. It is quite impossible, so far as I can see, to account for this change in intensity on either of them.

Experiments on Relative Phase (8). It can be shown by calculation or by graphic methods that when two pure tones of different wave-lengths are combined in different phases, the resulting air vibration varies greatly in shape. For example, if the tones have a ratio of 3 to 1 (that is, one tone being 426 d.v./sec. and the other one 142 d.v./sec.), it is found that their combination in one phase of relationship produces a curve with a plateau having a slight dip in the middle of it. On the contrary, their combination in a phase differing from the above by 180° produces a curve with a sharp peak followed by two slight undulations near the base line.

If the ear performs harmonic analysis on the composite wave which it receives, there should be no change in this analysis on changing the relative phases of the components. On the contrary, if harmonic analysis is not performed, it would be expected that some change would be produced in the sounds that are heard, since changing the phases produces such a marked effect on the sound vibrations which are reaching the ear. This can therefore be made the basis of a test to see whether harmonic analysis is performed by the ear or not. If it is performed, we should expect the observed sounds to be the same and to be independent of the relative phases in which the tones are reaching the ear. On the contrary, if no harmonic analysis is performed, we should expect some change to be observed. The experiment was performed in the following manner. Two tuning forks, both of them electrically driven, were taken. One of these was adjusted to vibrate at 426 d.v./sec., the other was adjusted to give 142 d.v./sec., i.e., in the ratio of 3 to 1; for the reason pointed out above, this ratio was selected in preference to others. These were driven by separate electrical circuits so as to prevent the possibility of one fork's being dragged into a particular phasal relationship with the other one. Wires were attached to the prongs of the forks; the other ends of the wires, which were three or four meters long, were attached to separate sounding boards, so that the tones produced by the forks were caused to combine near the observer's ear. The higher pitched fork sounded continuously, the lower one was alternately stopped and started at irregular intervals so that its phase should have a random relationship to that of the other fork. The experiment was repeated a number of times without any change in quality of the mixed tone being detectable. The inference is that the ear contains some apparatus by means of which harmonic analysis can be performed on the incoming sound vibrations. The only form of apparatus which could be accommodated within the dimensions of the internal ear is a system of resonators. This experiment is not only readily explained on the resonance hypothesis, but also a fundamental part of it appears to me to be inexplicable on any of the alternative hypotheses.

The Relationship between the Properties of Resonators. Experiment shows that in addition to natural period of vibration, there is another important property of the resonator, namely, the degree of damping (19, p. 143). This important factor controls the behavior of resonators in two ways: (1) the sharpness of their resonance, and (2) the persistence of their vibrations. Sharpness of resonance may be defined as the ability of a resonator to vibrate when its own tone and no other is incident on it. Persistence of vibration we can define as the number of vibrations it continues to perform after the forces which have set it into vibration have ceased to act. One or two examples may help to make these properties clear. Take, for instance, the diaphragm of a telephone. Its function is to vibrate with equal facility in response to sound waves of the whole audible range. Now it is clear that if it possessed any sharpness of resonance, it would respond by vibrating to certain selected tones and not to others. Such a property would make it very inefficient for the purpose in hand. Consequently, it is made as periodic as possible. This result is achieved by damping its natural vibrations by means of rubber pads. With regard to persistence of vibration, it is clear that its behavior will be spoiled if it continues in vibration after the sources of sound which have set it into vibration have ceased, because persistent vibration on its part would tend to fill up the intervals of silence which are so essential for clear speech. Its persistence should therefore be negligible, that is, it should be as dead beat as possible. This result is also achieved by padding it with rubber. It is clear, therefore, that the more highly damped the mechanical system is, the less selective does its resonance become and the less persistence it possesses. To take another example: An apparatus is required for releasing garage doors when the electric horn of the car which is kept in it is sounded outside. Such an apparatus should be made sharply tuned in order that it may release the doors when the particular tone is sounded, but not release them when other motor car horns sounding other tones chance to be sounded. A tuned piano string would be an example of such a sharply tuned resonator. Experiment shows that a difference of pitch of one or two vibrations per second will suffice to render the mechanism inoperative. That is, the garage doors will open to a musical note of 250 d.v./sec., and will not open to musical notes below 248 or above 252. With regard to persistence, such a piano wire when struck by its hammer will continue to vibrate with sufficient strength to emit an audible sound for one or two minutes. It is clear then in this case that sharpness of resonance is accompanied by the exhibition of marked persistence of vibration. A third example would be the tuning fork. This is so sharply tuned that under ordinary circumstances it requires the vibrations of another tuning fork to cause it to perform resonant vibrations. Its persistence is equally remarkable.

We see in all these cases that the increased sharpness of resonance is accompanied by increased persistence. This relationship Helmholtz (19, p. 143) indicated as in Table 2.

TABLE 2

Type	Sharpness of tuning	Persistence
V	⅛ tone	38 vibrations
W	¼ tone	19 vibrations
X	½ tone	9.5 vibrations
Y	1 tone	4.75 vibrations
Z	2 tones	2.37 vibrations

Helmholtz defined sharpness of tuning, in terms of the tone, as shown in column 2, in the following manner. Suppose a resonator to be set into resonant vibration by a tone of precisely its own natural frequency. Let this be 100 d.v./sec. for the sake of argument. By trying in turn other tones, e.g., 99 d.v./sec., 98.5 d.v./sec., and so on, that one is found which will cause it to vibrate with an amplitude which is precisely one-tenth that which it performs when the tone of its own natural frequency is sounded. Suppose this tone to be 99.5 d.v./sec. Then it is clear that it differs from the natural tone of the resonator by half a tone, that is, that it corresponds to X in the above table. The corresponding persistence will be seen to be 9.5 d.v. This means that, if it were set into vibration so that its amplitude was one-tenth of a millimeter, it would perform 9.5 d.v. before that amplitude was reduced to one-hundredth of a millimeter. A further 9.5 d.v. would be required to reduce the amplitude to one-thousandth of a millimeter, and so on. It is clear that the stretched string will correspond to the upper section of the above table, whereas the telephone, which has practically no sharpness of resonance and no persistence, would be placed on a line far below Z.

Now, if we demonstrate that the ear possesses the property of sharpness of resonance, we have shown that it has one of the properties which belong to resonators. If, secondly, we prove that the ear exhibits the phenomenon of persistence, we have shown that it possesses another property possessed by resonators. If, moreover, we can demonstrate that the sharpness coefficient and persistence coefficient correspond with those in Helmholtz' table above, we have produced well-nigh irrefutable evidence that there are resonators in the internal ear. The following experimental method used by Mayer (26) will enable us to measure the sharpness of resonance of the ear. Two tuning forks of nearly the same pitch are taken. One of them can be made sharper or flatter than the other by any required amount. When they are simultaneously struck, audible beats are produced. These beats decrease in intensity and increase in frequency as the difference between the pitches of the two tuning forks is made greater. By careful adjustment the forks are arranged so that the beats have just become inaudible. The difference of pitch between the two forks is now carefully measured. It defines the sharpness of tuning of the ear, because beats will be produced only when both tones set into appreciable vibration the same area of basilar membrane. If the sharpness of tuning of the basilar membrane is approximately the same for each of these tones, this area must be situated midway between them, and may be quite small. When the tones have been so far separated that the beats become inaudible, they are no longer simultaneously affecting this intervening area, and so we can say that their sharpness of resonance is represented by one-half the difference in frequency between them. For example, suppose one tone to have been 250 d.v./ per sec., and the other tone to have been 230 d.v./ per sec., then the sharpness of resonance is represented by 10 tones. This number of vibrations represents about one-third of a tone. Therefore, the ear would exhibit roughly the same behavior as resonators between W and X in the above table.

The following experimental methods have been used to measure the persistence of the ear: (*a*) the number of trills or shakes which can be clearly heard by the ear (18); (*b*) the speed at which a disc must be rotated so as to just produce inaudible pulsations in the musical tone which is being interrupted. The experiments of Allen and Weinberg have already been alluded to above (42). The experiments of Mayer (26, p. 18) must also be alluded to. It is quite clear from all these experiments that some mechanism in the ear continues in vibration after the musical tone which sets up those vibrations has ceased to sound. In other words, the ear exhibits one of the characteristic properties of resonators. One example may be given. At middle C (256 d.v./sec.), Mayer found that the ear did not suffer an appreciable diminution in the intensity of sound unless a silent interval corresponding to a longer time than two vibrations was allowed to elapse.

Figures will now be given in support of the statement that the sharpness of resonance coefficient and the persistence of vibration coefficient for tones of different pitch bear the relationship to be expected according to Helmholtz' table. These figures are given in Table 3. If the S and P columns of Helmholtz' table given above be multiplied together, the product is in each case 4.75. When the S and P factors in the above table are multiplied together, it will be seen that they average about 21. The reason why they do not give the same product as those of Helmholtz' table is because Helmholtz calculated his figures on the assumption that the amplitude of vibration was reduced to one-tenth

TABLE 3

Mean tone	Sharpness factor	Persistence factor	S × P
128	12.70	1.78	22.60
256	10.00	2.06	20.60
320	9.45	2.19	20.70
384	9.07	2.18	19.75
512	8.45	2.37	20.00
640	8.15	2.54	20.65
760	7.82	2.68*	20.95
1024	7.22	3.01	21.70

*The mean tone for this persistence factor was actually 768. The fourth column figures (S × P) are obtained by multiplying the sharpness factor by the persistence factor.

its initial value in the case of both coefficients. On the other hand, the values found for the ear depend on the sensitivity of the hair-cells. Their absolute threshold is involved in the measurement of the sharpness factor, whereas their difference threshold is involved in determining the persistence factor. The values of neither of these thresholds are accurately known, although Hartridge's calculations lead him to suppose that the absolute threshold has a value of about 6%, while the difference threshold has a value of about 2.5%. In consequence, neither the sharpness factor nor persistence factor is identical with Helmholtz' sharpness coefficient and persistence coefficient. In consequence, their products are different. But the point of importance is that the products in the above table are nearly constant, as they are on Helmholtz' table, and this would be the case only if the ear contained a system of resonators.

To summarize, resonators exhibit three properties: (*a*) sharpness of resonance; (*b*) persistence of vibration; and (*c*) a constant product for these two factors. The ear shows the existence of these three properties. The conclusion is that it contains resonators.

V. Reply to Criticisms

Henson discovered that certain tones set hairs of the opossum shrimp into vibration and other tones affected other hairs. If a system of resonators can be established so simply, Keith (30, p. 188) asks why so complex a structure as the organ of Corti should have been elaborated for this purpose.

The reply is that in the opossum shrimp the hairs are free. If we could connect them together by a membrane as is done in the basilar membrane, a more sensitive arrangement would be obtained for picking up external sounds; and if further we could mount the whole apparatus in a tube filled with liquid and could set that liquid in motion by means of a piston which was itself actuated by a mechanism which provided a mechanical advantage of about twenty times, we should find a still more effective organ of hearing resulted. So far as can be judged, the ear has been designed in such a way as to make it sensitive to sounds of very low intensity.

Keith (21, p. 174) has objected to the resonance theory on the ground that so wide a range of musical tones could not be received by resonators which vary in length by so small an amount.

In Table 4 are given the greatest and least lengths for basilar membrane fibers as measured by different observers. It will be seen that there is a

TABLE 4

Observer	Greatest	Least	Ratio
Keith	.300	.120	2.5 to 1
Retzius	.240	.140	1.7 to 1
Henle	.410	.023	17.8 to 1
Henson*	.370*	.030*	12.5 to 1

*Inferred.

big variation in the estimates which have been made. Let us assume for the sake of argument that Keith's values represent the true ones, and that the number of vibrations per second performed by the short fibers is 4000 and the longest ones 40. Then, applying the method which Roughton (18, p. 365) suggested to Hartridge, the following formula is made use of:

$$\frac{T}{A} = 4R^2 \times L^2$$

We find then that the tension in grams per square mm. equals 9.2 for the shortest fiber and .00006 for the longest. Now the higher value is negligibly small compared with the stress which animal fibers will bear without breaking. Thus, tendon will support 23,000 grams per square mm. We have no difficulty in supposing, therefore, that the basilar membrane fibers will readily support the load to which they would be subjected according to the resonance theory.

Meyer (27, p. 19) has advanced the objection "that theories which compare the fibers of the basilar membrane to resonators introduce in opposition to the laws of biology the idea of a permanent constant tension of the basilar membrane. Authors who expect their readers to accept it as a truth should first of all try to convince them of the possibility of living animal tissues retaining their tension for any length of time."

The reply to the above criticism is that many living tissues are known which do, in point of fact, continue under tension during the greater part of a man's life: (*a*) the heart contracts periodically without cessation; (*b*) the sphincters of bladder and anus are nearly continuously in contraction; (*c*) the diaphragm is in constant tension; (*d*) the coats of the eye constantly withstand the intra-ocular tension; (*e*) the muscles are nearly in constant tone; (*f*) the ligamentum nuchae of the horse, etc., support the weight of the head nearly continuously. The conclusion from the above examples is that living tissues do retain their tensions for very long periods of time. Is there any reason why the basilar membrane fibers should not do so too?

Keith (21, p. 187) has objected to the resonance theory on the ground that "there are no anatomical structures which can serve as resonators in the cochlea." "The basilar fibers are not free strings, they are clothed on their lower surface by a layer of loose cells and fibers, and their upper surface is clothed by a stratum of cells."

The answer to this criticism has already been given in a previous section. The cells play, so far as can be judged, an essential part since they load the fibers of the basilar membrane and therefore cause them to vibrate to lower tones than they otherwise would do. When the cells are examined in detail, it is seen that the short fibers which respond to high tones are only lightly loaded with cells, whereas the long fibers are heavily loaded so as to cause them to vibrate to the lower tones of the audible scale.

Wrightson (47, p. 104) objected to the resonance theory on the grounds that the fibers of the basilar membrane are not free but are connected sideways.

Now calculation shows that the sharpness of resonance of the ear resonators near middle C is approximately 0.6 of a tone. That is, it is slightly worse than Type X in Helmholtz' table. Now the steepest part of the curve which represents this sharpness of resonance has a slope which corresponds to a 5% variation of amplitude for each one-hundredth of a tone variation in pitch. Now the maximum number of separate resonators in the human ear is, according to Keith, about 4000. Walduyer has given the number as 4500. If we say that the musical scale extends from 64 vibrations per second to 4100, i.e., over about six octaves, and if we assume that 600 resonators are devoted to each of these octaves, then 3600 resonators will have been accounted for, leaving about 400 resonators to deal with the lower limit from, say, 16 to 64 d.v./sec., and the upper limit from, say, 4100 to 40,000. This number would appear to be more than sufficient. McKendrick estimated that about 110 would suffice (35, Vol. II, p. 1184). If we assume then that 600 resonators are devoted to each of the octaves within the normal musical range, and remembering that there are six whole tones in each octave, approximately 100 resonators will be represented in the interval of one tone. That is, between one resonator and its next-door neighbor there will be a difference of pitch of approximately one-hundredth of a tone.

Now Hartridge has calculated above that a variation of amplitude of 5% is the maximum met with for each one-hundredth of a tone variation in pitch. The conclusion is, therefore, that the greatest possible difference of amplitude performed by any two neighboring fibers of the basilar membrane will be one of 5%. The bands which connect one fiber with the next, which are agreed by all histologists to be of a lax description, should be readily able to permit such a difference of amplitude.

Kolmer found that one single fiber of the auditory nerve may by branching supply two separate hair-cells. Held concluded that a single nerve-fiber may be distributed over a quarter turn of the cochlea. On the contrary, Snodgrass found that in an adult with normal hearing, there were approximately 14,000 fibers in the cochlear division of the auditory nerve. Keith states that the number is 15,000. Now McKendrick's estimate of the number of hair-cells in man is a little over 15,000, so that practically each hair-cell has its own separate nerve-fiber and very few share nerve-fibers. Suppose, for the sake of argument, careful histological examination demonstrated that the three hair-cells which lie on a particular section of the basilar membrane were connected to one and the same nerve-fiber; it would indicate only that this particular section is less efficient than it might otherwise be. Such a state of affairs might be found, for example, at the upper and lower ends of the cochlea without appreciable loss of efficiency. In much the same way, the rods in the periphery of the retina are frequently connected in small groups to one and the same optic nerve-fiber. In any case, visual acuity is low because the images produced by the optical system of the eye suffer from rather serious defects at the periphery. In the case of the ear, the recognition of musical tones of the upper and lower parts of the range of audition is known to be distinctly below that found nearer the center of the range.

Luciani (24, p. 236) objects that "phylogenetically the simultaneous development of the excessively numerous specific differentiations of the fibers of the auditory nerve offers great difficulty."

The reply to the above criticism can be obtained by analogy with the eye. We find these varying in development from the simplest to the most complex, and the matter presents no difficulties to our minds. In the case of the ear, also, we can trace its development from simple beginnings through the frog, alligator, and bird, to reach a high state of development in mammals. It is true that the auditory nerve contains a very great many fibers, but the number is small compared with that of the optic nerve.

Keith (21, p. 60) states that even on Helmholtz' hypothesis "the final process of analysis must be done in the brain . . . because the central exchange has only one key or clue to the particular station or stations in which the messages come, namely, the rate at which the stimuli or messages arrive."

Keith stated on the previous page that "Young had postulated three different kinds of nerves to the retina; von Helmholtz went a considerable degree further and postulated many thousands of different kinds of nerves for the ear. The most impossible aspect of von Helmholtz' theory comes to light when we try to carry it beyond the cochlea into the nervous system." The reply to the above criticism is that in just the same way that we learn by experience the relative positions of our fingers, so we learn the relative positions of the different hair-cells of our cochleae. It is true that we cannot run our fingers up and down our basilar membranes as we would on the keyboard of a piano and so correlate touch with hearing. But neither can we do this with our retinas, and yet we very quickly learn how the different rods and cones are positioned with regard to one another. Is it not possible that different combinations of sound waves can educate the hair-cells of our cochleae in the same way that different patterns of light waves teach our eyes? It is most unlikely that the higher centers of the brain use the rate at which the stimuli arrive as the clue to the pitch of the tone which is being received, any more than the eye appreciates color by transmitting to the calcarine cortex nerve-impulses having the frequencies of light.

VI. Watt's Hypothesis

Watt's theory of hearing (41) was devised as "an episode in a larger effort." It was his "firm belief that all our sensory experiences can be completely accounted for in general systematic terms without our having recourse to the discoveries of physics or physiology at all." He wished to build up a psychology of experience which should be built out of "psychical stuff" and "to treat [this stuff] as a real object, having properties we can only discover through careful cognitive investigations."

Analyzing our sensation we have, according to Watt, six attributes: quality; intensity; duration; "order," systemic and temporal; and extensity. In hearing, the temporal order gives the succession of sounds, while the systemic ordinal attribute gives us pitch, for, phenomenally, pitches follow one another. They are given names in order, e.g., *a, b, c, d,* etc.; one pitch is said to be higher than another, and so on. But pitches, besides rising and falling, vary in volume or mass. "Low tones are great and massive and all-pervasive, high tones are sharp, thin, circumscribed; as tones rise in pitch, their volume sinks gradually, pulling itself together as it were, till it is finally almost too small and thin to be noticed." This volume or mass is not physical volume or mass—it is "tonal volume, an experiential volume. This volume is the attribute of extensity. Although volume varies with pitch, being greater the lower the pitch, discrimination of the latter attribute is the finer. The other attributes give us loudness (intensity), tones, noises, etc. (quality), and duration. Since sounds possess volume, tones are not atoms of auditory sensation—they are rather molecules, the smallest masses we find. The pitch is the intensively predominant atom of the whole volume. The volume in a pure tone is symmetrically balanced in relation to pitch. For example, a pure tone can be represented as in Figure 1 where Vh—Vl represents the total volume of the tone, Vl being the lower extreme and Vh the higher. P represents the predominant (systemic order) that constitutes the pitch, while the perpendicular Pi represents the relative intensity in which that predominant order is present." The intensities of the other orders, such as X, are represented by the perpendiculars at those orders. All tones, lower in pitch and having smaller volumes than this tone, fall within the volume Vh–Vl, for "tones played together always seem together, inseparate, blended or fused, no matter how different their pitches may be." Also pitch moves progressively to one side of a series, as we listen to the tones of the piano played from one end to the other of the keyboard. The tones of higher pitch will therefore all be crowded within the volume Vh—Vl, and towards the end Vh. When octaves are played together, the volume of the upper tone will lie within the upper half of the volume of the lower. Since it will have a point of intensity at its middle, the total volume of the two tones will be the same as that of the lower tone in the lower half, while in the upper half the volume will be the sum of the two volumes with a second prominence at P_2 which gives the pitch of the upper tone, as shown by the full line in Figure 2. For the two tones of the fifth, c and g, the volumes will be as indicated in Figure 3—the upper limits of the two volumes will be identical at Vh; the lower limits will be Vl_c and Vl_g where the distance

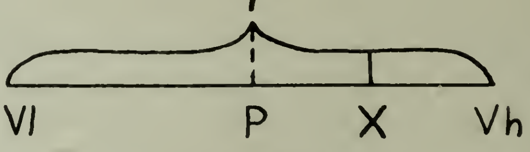


FIGURE 1
Representation of a Pure Tone

$Vh—Vl_g = 2/3\ (Vh—Vl_c)$. Similarly the pitches will be at P_c and P_g where $Vh—P_g = 2/3\ (Vh—P_c)$. The resultant volume will be as indicated by the full line—"so that harmonies and the 'musical colour' of tones are like an ornamentation of the outline of the volume or extent of the lowest tone that is played or heard." When listening to music, the ear follows the changes in the prominences and psychical volumes as they appear. Hence, the degree of tonal fusion of intervals such as octaves, fifths, etc., depends on the symmetry and balance of the volumic outlines. According to one's training and interests, these may be followed in greater or lesser detail, and pleasure will be derived if the changes satisfy one's aesthetic tastes.

"Noises are essentially the same things as tones, only tones are regular and of determinate pitch, while noises are irregular and of indeterminate pitch. Not that they have no pitch at all. It merely changes rapidly, or there are many pitches at once, so that one predominates so as to determine which pitch the mass of sound shall be held to have." A noise may

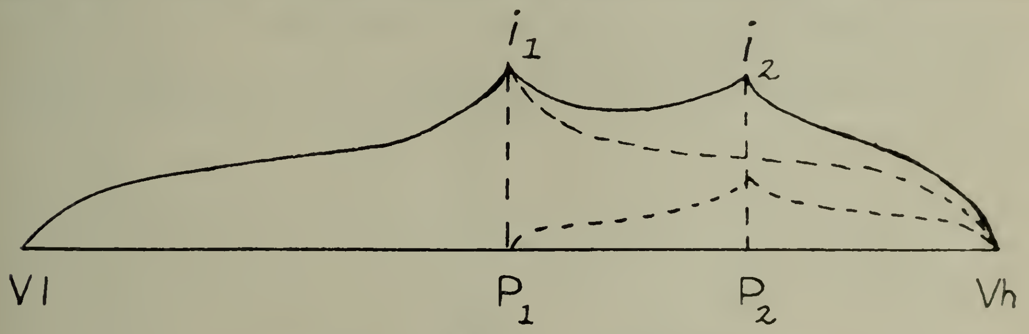


FIGURE 2

Representation of a Pure Tone and Its Octave

(From: Watt, H. J. The psychology of sound. Cambridge: Cambridge Univ. Press, 1917. P. 70.)

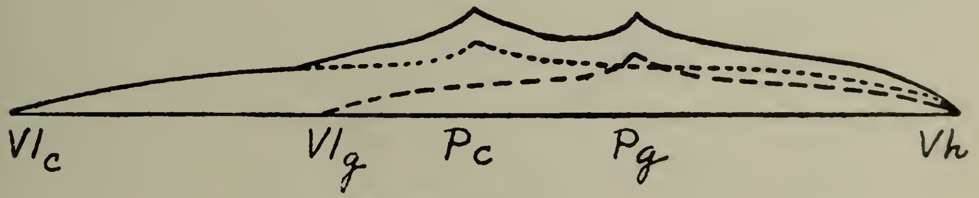


FIGURE 3

Representation of a Pure Tone and Its Fifth

(From: Watt, H. J. The psychology of sound. Cambridge: Cambridge Univ. Press, 1917. P. 70.)

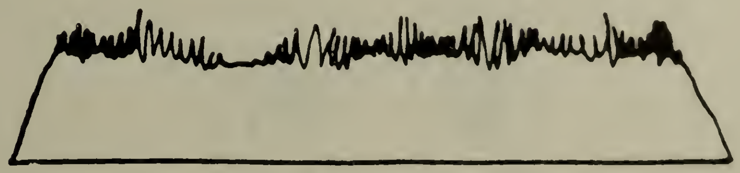


FIGURE 4

Representation of a Noise

then be represented by some such diagram as Figure 4. The whole of this is the basis of the psychology of hearing, according to Watt. An examination of the various physiological theories of hearing leads Watt to the acceptance with certain modifications of the advancing wave type of theory, as formulated by Kuile. When a sound wave falls on the ear, the stapes is made to vibrate. The pressure exerted on the fluid in the cochlea by an inward movement of the stapes produces a bulge in the basilar membrane which begins at the base of the membrane and reaches a distance along it which is proportional to the wave-length of the sound wave. This bulge has a point of maximal bulge approximately in the center of the length affected. It is this length of the basilar membrane, affected by any given sound, which determines its volume; the point or points of the basilar membrane which are maximally affected determine the pitch or pitches of the tone, while the amplitude of the displacement of the basilar membrane determines the loudness. On the physical side it is wave-length which determines the psychical "volume," the position of the maximum condensation which determines pitch, and the amplitude of the (air-borne) vibration which determines loudness.

There is much in this theory that appeals to the psychologist. In at-

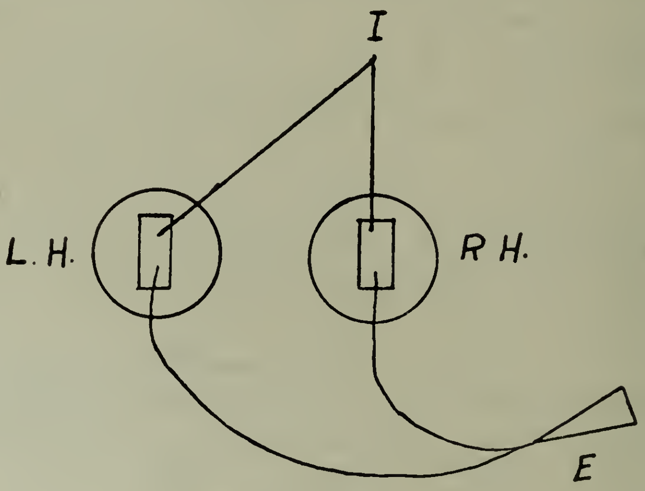


FIGURE 5
E = Ear
R. H. = Right Hemisphere
L. H. = Left Hemisphere
I = Integration

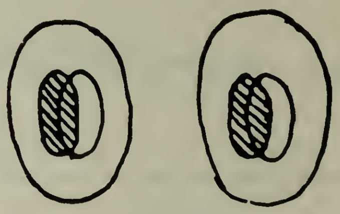


FIGURE 6

tempting to formulate a theory of the action of any sense-organ, one may begin with a study of the organ itself, but the theory so deduced must stand or fall according as it fits in with the facts of experience. Watt begins with a study of the phenomena and accepts those physiological findings which fit in with the psychological facts. With this method we agree —but we must examine his facts. If these are incorrect, his theory falls to the ground, without its being necessary in any way to appeal to the results of physiological research. His theory falls if volume is not a fundamental attribute. This, as has been shown in the previous chapter (see section on Volume), is our opinion.

Another objection to the theory is that the basilar membrane cannot vibrate in the way the theory requires, and that loudness is not proportional to the amplitude of the vibration, but to the square of the amplitude (22, 23).

VII. Boring's Hypothesis

This hypothesis,[1] like Watt's, is based on the phenomenological side of hearing, chiefly in connection with a theoretical physiology of the cortex.

The nervous impulses aroused in the auditory nerve-fibers of the organ of Corti stimulate the auditory areas in the hemispheres—both hemispheres being excited by the impulses from each ear. If *E* (Figure 5) represents the right ear (the triangle representing the basilar membrane), then the nervous impulses pass along to the right and left hemispheres (*R.H.* and *L.H.*), exciting a certain area in each, the size of the area depending on the length of the basilar membrane which is disturbed sufficiently for the nervous impulse to be initiated.[2]

The length of the basilar membrane which is disturbed depends directly, Boring assumes, on the amplitude of the vibration of the sound wave, so that there is a direct linear relation between amplitude of vibration and number of cells in the cortex which is excited. The volume of the sound depends on the magnitude of the cortical area which is excited.

The integration[3] of these excitations—schematically represented at *I* in the diagram[4]—gives the cerebral correlation of loudness which also varies as the amplitude of the vibration, not as the energy of the sound wave. Both volume and loudness depend on the amplitude of the sound wave, but, having different limen which vary according to different laws, they are not the same, so that some integration to give loudness is necessary.

---

[1] Professor Boring considers that this hypothesis explains the facts of audition better than other theories, but does not claim that it is correct. For purposes of exposition it is assumed that it is a hypothesis he would uphold.

[2] The diagrams are not given by Professor Boring. I have drawn them in the attempt to set out the hypothesis more lucidly.

[3] "On the whole it is plausible, even if not essential, to suppose some kind of integration in the brain where multiplicity of impulses is translated back into degree of excitation" (6, p. 161).

[4] It is not suggested that the integration takes place at some other point. *I* is shown separately to indicate that this integration to give loudness is another process, separate from but dependent upon the excitation of the cells of the cortex.

Pitch is, on this theory, dependent on the frequency of the impulses transmitted along the nerve-fibers.

The relations between pitch, intensity, and volume follow: Loud tones will be more voluminous than soft tones; high-pitched sounds will be smaller in volume than low ones having the same physical intensity, for, since they vibrate more rapidly, they will have the smaller amplitude. Auditory localization—the hearing of a sound on the right if it is louder, or if the waves are earlier in time or phase, on that side—is explained by assuming that the two cortical areas in each hemisphere which are excited by the two ears are adjacent regions which are very close together and in part overlap (Figure 6, where the shaded area corresponds to the region excited by the left ear, and the unshaded to that excited by the right, or vice versa). If the one area is more intensely stimulated, the mean point of stimulation will be to that side, thus giving localization by intensity; localization by phase difference and by time difference is to be explained by the inhibition of the later by the prior excitation, giving a mean position to the side leading in phase or time.

The theory has been attacked on various grounds: (*a*) that the attribute of volume is not a primary attribute (4); (*b*) that nerve-fibers cannot transmit vibrations at the rate of 20,000 or more per second (44); and (*c*) that the theory fails entirely to explain the facts of sound localization (4). In addition, (*d*) Boring himself confesses the inability of the theory to explain the fact that, when a complex tone is heard, the trained ear may discriminate the various pure tones of which the complex tone is built up.

VIII. The Resonance Hypothesis

It is possible that one reason why the resonance theory meets with objection is that there is great difficulty in picturing the events taking place in a large number of resonators of different pitches when musical tones commence to fall on the ear and when these same musical tones cease. It is therefore proposed to describe as concisely as possible what one would observe taking place according to the resonance theory if one could watch the cochlea at work. Let us suppose the cochlea to be at rest and that a pure musical tone of constant intensity of 254 d.v./sec. commences to fall on the ear. Then it is clear that the sound-wave energy entering the ear in unit time remains constant for as long a time as the musical tone continues, because that tone is constant in pitch and amplitude. In consequence, the total energy to be distributed between the different fibers composing the cochlea must be directly proportional to the length of time during which the musical tone has been arriving, if we assume either that no energy is lost in overcoming friction or producing eddies or if these losses are constant. On this basis it is possible to calculate the number of resonators which will be in vibration after any given number of vibrations and the average amplitude of vibration of these same resonators. It is also possible to calculate the amplitude of vibrations of the resonator which is strictly in tune with the incoming musical tone, that is, 254 d.v./sec.

TABLE 5

Number of sound waves from commencement of tone	Number of resonators in vibration	Average amplitude of vibration of reasonators in percentages	Percentage amplitude of "in-tune" resonator
1	12,000	.3	3
2	2,000	3.0	9
3	1,200	8.0	17
4	900	13.0	27
5	700	21.0	39
6	580	31.0	50
7	480	44.0	61
8	400	60.0	72
9	340	80.0	83
10	300	90.0	93
∞	240	100.0	100

In calculating Table 5, it has been assumed that the sharpness of resonance of the "in-tune" resonator is 0.6 of a tone on Helmholtz' scale. It will be seen that after the arrival of only one sound vibration, 12,000 resonators are vibrating with an average amplitude of 0.3%. If no previous sound waves have fallen on the ear for some little time, it is quite possible that this amplitude of vibration will stimulate the auditory hair-cells. This will cause a large number of auditory nerve-fibers to send impulses up to the temporal area of the cerebrum. On the other hand, if the ear has been recently receiving previous tones, there is experimental evidence which suggests that any intensity less than 3% of the maximum remains unnoticed because it is below the threshold of stimulation of the hair-cells. In this case, then, we see that the "in-tune" resonator will be vibrating with an amplitude of 3%, i.e., its swings will just have sufficient amplitude to cause stimulation of the hair-cells attached to it. The rest of the resonators in vibration have an amplitude which is, on the average, one-tenth this, according to the above table, and therefore their vibrations will be below the threshold value and their effects will therefore not be transmitted to higher centers. After two vibrations have been received by the ear, 2000 resonators are calculated to have an average amplitude of 3%, which is the threshold value. On the other hand, the "in-tune" resonator which forms one of these has reached an amplitude of 9% which is three times the threshold value. Consequently, certain of these 2000 must have an amplitude which is below the average, and therefore the area of the cochlea from which impulses are travelling up to higher centers must be less than 2000. An inspection of the table will show that if the incoming sound vibrations continue to affect the cochlear mechanism, the number of resonators rapidly decreases until, after ten vibrations have been received, the amplitude of the "in-tune" resonator is 93% of the value it will finally reach if the sound waves continue to arrive for an infinite length of time. On the other hand, after seven vibrations have been received, the number of resonators in vibration is roughly only twice the number which will finally be remaining in vibration when the sound continues to arrive for an infinite time. At the commencement of a musical tone, we should, therefore, expect to find (*a*) an element of uncertainty in the pitch since the "in-tune" resonator has an amplitude which is not markedly better than that of surrounding resonators, and (*b*) the sensation of impurity in the musical tone owing to the fact that a wider range of resonators is temporarily set in vibration compared with those which normally respond to a pure tone. With regard to the first point: This deduction fits the experiments of Kohlrausch (31, p. 452) in which a musical tone of short duration was compared with a musical source of sound of measurable pitch. A comparison was made between them to see how much sharper or flatter the variable tone had to be for the difference to be noticed. It was found that after sixteen variations the pitches had to be within 1% of one another, whereas for a musical tone of three vibrations an accuracy of 2% sufficed, and for a musical tone of two vibrations' duration 3% accuracy was sufficient. The above experiment shows that a musical tone of very short duration has a somewhat indefinite pitch. With regard to the second point: The impurity of sensation at the commencement of a tone is a factor of fairly constant recognition. A number of musical instruments can be recognized from the kind of roughness which is produced at the commencement of their tones, e.g., the guitar, mandolin, banjo, and various other examples of stringed instruments where plucking or striking of the strings takes place. It can be demonstrated very simply by causing a musical tone to undergo a short interruption. This fact will be referred to later. We are now in a position to consider how soon after the commencement of a tone will it have apparently reached the steady state, with the disappearance of the phenomena that, as we have shown above, take place at the beginning of the musical tone. On reference to Table 5, it will be seen that the "in-tune" resonator has nearly reached its maximum amplitude and that the number of resonators which accompany it on either side is only 20% greater than the final number after ten vibrations of the incoming sound have reached the ear. Now the musical tone for which these figures were calculated was 254 d.v./per sec., and therefore ten vibrations of this tone will occur in approximately one–twenty-fifth of a second. This is therefore the time required for the ear to reach very nearly the steady state.

We must now consider how exactly the higher centers perceive this musical tone and differentiate it from others. In Table 6 is shown the amplitude performed by the different resonators which lie to either side of the "in-tune" one.

No. 0 in Table 6 represents the "in-tune" resonator. No. 1 represents the resonator immediately next it on both the low-pitched and high-pitched sides. No. 2 represents the resonator immediately next both No. 1 resonators, and so on. It will be seen that a line is drawn between resonator No. 2 and No. 3. Calculation shows that the difference threshold for intensity is past where this line occurs. A similar line will be seen following resonator No. 80. The amplitude below this point falls under the absolute threshold of intensity of the ear. Therefore approximately

TABLE 6

Number	Amplitude	Number	Amplitude
0	100	9	86
1	99.3	12	74
2	98	18	55
3	97	24	41
4	96	36	24
5	94.7	48	15
6	93	60	10
7	91	80	6

80 resonators on either side of the "in-tune" one are sending up impulses along their corresponding auditory nerve-fibers, that is, 161 in all. In the first place, I should like to point out how very different this number is from the usual assumption that one or two or, at most, half a dozen resonators on either side of the "in-tune" one are performing sufficiently intense vibrations for their hair-cells to be affected. It should be observed that these resonators on either side of the "in-tune" one which are accompanying its vibration are not vibrating at the rate which corresponds to their own natural frequency. On the contrary, both those of longer pitch and those of shorter pitch are vibrating at precisely the same rate of vibration as the "in-tune" resonator. That is, the whole area of basilar membrane affected moves up and down at one and the same time. Now the question has frequently been asked how, in the first place, these different resonators could vibrate each according to its own natural period, seeing that they are all connected by lax fibers. The reply is that the lax fibers do not come into the question, for even if all the fibers were free they would all perform forced vibrations in step with the one strictly in tune with the musical tone. The other question which has frequently been asked is why, if such a wide band of resonators be set into vibration, the ear hears a pure tone. Would it not be more reasonable to expect the ear to hear a markedly discordant tone since the 160 resonators which are in vibration correspond to nearly 1.6/ tones? The reply to this question is that this state of affairs is the usual one. It always has been the case ever since the cochlea first functioned, and presumably it will continue to be the case unless some pathological change takes place. When we look at a distant point source of light we cause an optical image to fall on our retinas which has dimensions far greater than those of any one retinal cone. It can be proved conclusively that a considerable retinal area receives adequate stimulation. This does not cause discordant results. It is the habitual state of affairs. If we prick ourselves with a needle, a considerable number of sensory nerve-endings, possibly of various kinds, receive stimulation, but discordant sensations do not result. It should be noted that the figures in the above table are at variance with a suggestion put forward by Gray (14) which he has called "the principle of maximum stimulation." The difference in amplitude between the strictly "in-tune" resonator and the

ones which flank it on either side is so small that it falls below what I believe is the value of the smallest perceptible difference of amplitude which, as I have pointed out above, has approximately the value of 3%. So far as can be seen not only the "in-tune" resonator but, in addition, the two which flank it on either side, which make five altogether, vibrate with so nearly the same amplitude that they send identically the same intensity of impulse up their respective auditory nerve-fibers. The principle of maximum stimulation could thus be applied only if five resonating elements could be regarded as sending impulses of a different amount up their respective nerve-fibers. The resonators which lie outside this central group of five are performing vibrations which are markedly less, so that the impulses which are sent up to their respective auditory nerve-fibers are physiologically smaller than those sent up from the central group. In consequence, the strengths of impulses gradually fall off as the more peripheral part of the area of basilar membrane is considered, and this gradual decrease goes on until the 80th resonator away from the "in-tune" one is reached. at which point the absolute threshold is reached and no further stimulation of the auditory nerve-fibers occurs. In just the same way that we regard the optical image of a bright source of light to consist of an area having a center with a peripheral margin which gradually decreases from bright to dark as one passes away from the center towards the extreme edge, so we also regard the sound image of a pure tone to consist of an area having an intense center with a peripheral margin which gradually decreases from loud to soft as one passes away from the center towards the extreme edges.

We must now briefly consider the events taking place in the cochlea on the cessation of a pure tone. Table 7 shows the amplitudes of all the resonators which have been in vibration during the time that those vibrations die away to silence. The line above No. 1 shows that the amplitude of resonant vibration fell almost immediately after the exciting sound waves stopped below the difference threshold of stimulation. The line between No. 9 and No. 10 indicates that at about the amplitude corresponding to this number of vibrations the intensity of sound fell below the absolute threshold. In the case of the "in-tune" resonator, therefore, almost immediately after the exciting tone has stopped, the difference threshold is

TABLE 7

Vibrations	Amplitude	Vibrations	Amplitude
1	74	9	7.4
2	56	10	5.6
3	42	11	4.2
4	32	12	3.2
5	22	13	2.2
6	16	14	1.6
7	13	15	1.3
8	10	16	1.0

passed, and therefore impulses passing up the auditory nerve-fibers should inform them that some marked diminution of intensity has taken place. At the end of nine vibrations, the absolute threshold is passed and the auditory nerve-fibers cease to send impulses to the higher centers. Now these values have been calculated for a tone of 256 d.v./sec., nine vibrations will correspond to a time interval of approximately one–twenty-eighth part of a second. It should be noted that when the exciting tone ceases, the resonators on either side of the "in-tune" tone, which have hitherto been vibrating with a frequency which was different from their own natural period of vibration, are no longer constrained to do this. When the exciting tone ceases, they are free to continue to vibrate for the short time indicated by the above table according to their own natural period of vibration. The higher pitched ones will therefore come to rest a very short time in front of the lower pitched ones.

With regard to other tones in the musical scale above and below 256 d.v./sec., Table 8 shows roughly the time taken for the musical tone to begin. In the last column is shown the approximate time taken by its termination. The values in the second column for sharpness of resonance are in decimals of a tone, as in Helmholtz' Table 1. It will be seen that both beginning and finishing take longer for tones of low pitch than they do for high ones. This fits in with the well-known fact that more rapid shakes can be made in the treble and still remain clearly audible than they can in the base. It will also be seen that low notes suffer with respect to sharpness of resonance also. This may be correlated with the fact that the pitch of low tones is more indefinite than is the pitch of high tones, and that, moreover, beats occur between musical intervals in the base which are free from beats in the treble.

TABLE 8

Tone	Sharpness of resonance	Persistence of vibration	Time taken to begin	Time taken to finish
128	0.73	6.5	0.063	0.057
256	0.59	8.0	0.040	0.036
512	0.50	9.5	0.023	0.021
1024	0.43	11.0	0.011	0.012

Brief consideration must be given to silent intervals of various durations. If the interval is greater than one-tenth of a second, then musical tones of all pitches have time to die away to complete silence. If, however, the interval is considerably shorter than this, the dying-away only partially takes place. For example, if a tone of 256 d.v./sec. is interrupted for two vibrations, then the amplitude of vibration has time to fall to not more than half its initial value, as will be seen from Table 7. In consequence of this, another factor may come in which will partially mask the decrease in amplitude, namely, the vibrations set up in outlying parts of the basilar membrane by the recommencement of the tone. Thus, under certain circumstances instead of a brief silence, a brief noise is heard owing to the temporary vibrations of the outlying resonators.

There is clear experimental evidence not only that the tones of different instruments vary in respect to the partials (overtones) which accompany the fundamental tone and in the manner of their commencement and termination, but also that the different sounds produced by the human voice differ in a similar manner. The recognition of partials on the resonance theory is as simple as the recognition of the fundamental tones themselves. Each partial is perceived in the same way as it would be if it were the only tone setting the basilar membrane into vibration. Even if proof were definitely obtained that the partials in the case of the human voice bore a relationship to the fundamental tone which was of an inharmonic character, this would not matter so far as the resonance theory is concerned.

The consideration has so far been given to pure tones of constant pitch. The question therefore arises how will the behavior of the cochlea be affected if the tone which is received undergoes variations in pitch. Let us briefly consider two different cases: (*a*) where the tone varies periodically in pitch, first rising and then falling in a uniform manner, as in the *tremulo* performed by a violinist; (*b*) where the tone rises or falls continuously in pitch, e.g., the sound made by a siren while it is starting or stopping. This is called *portamento.* In (*a*), if the variation in pitch takes place at a slower speed than about 10 times per second, the area of basilar membrane affected is but slightly larger than that which would be affected by a pure tone of the same pitch. This area moves up and down the cochlea, first towards higher pitched tones and then towards lower ones at the same rate as the musical tone does, but with a slight time lag. If, on the other hand, the rate of change is much more rapid than 10 times per second, there would seem to be good reason to suppose that the area of basilar membrane affected would be distinctly greater than that affected by a single pure tone, because, for example, the resonators affected by the tone when it has risen above its normal pitch have not time to come to rest during its passage to lower pitches and its return. In addition, it would seem likely that the variations in amplitude of the "in-tune" resonators take a greater share in conveying to the higher centers the pitch variations which are taking place than do the amplitude variations of the more outlying ones because, the amplitude of the former being greater than the latter, the fall of amplitude which takes place with the change of pitch is correspondingly greater. With regard to portamento, experiments were performed with a model consisting of a number of resonators supported by a movable beam. This was driven by a swinging pendulum which could be varied in length. Observation showed that the resonant vibrations die out on one side of the area of basilar membrane which is in vibration, while at the same time it tends to be initiated on the other side. This causes the vibration to transfer itself steadily either to higher or lower tones according as the incoming musical tone is rising or falling in pitch. Calculation indicates that the area of basilar membrane in vibration at any moment is somewhat larger than that affected by a stationary pure tone. If the pitch of the tone changes by as much as an octave per second, the group of resonators affected would appear to be as much as two or three times as large as that sent into resonant vibration by a pure tone. We conclude that the tone will be readily recognized but that it will not have the same purity as a fixed tone possesses. This latter effect may possibly be correlated with the unpleasant character of such tones, e.g., the commencement of a blast of a steam siren.

The way that tones of different intensity are perceived by the cochlea, according to the resonance theory, will now be dealt with. A tone of very low intensity indeed will set an area of basilar membrane into resonant vibration with amplitudes which are shown in Table 6. The amplitude of vibration even of the "in-tune" resonator will, however, be extremely small. In consequence, only the vibrations of this resonator and perhaps a few others on either side of it will stimulate the hair-cells with sufficient intensity to be above their absolute thresholds of stimulation. Consequently a few fibers of the auditory nerve will convey impulses of an all-or-none nature to the higher centers, and in addition these impulses will have considerable intervals of time between them. If this same tone slowly increases in intensity, the all-or-none impulses conveyed by these nerves will travel at a more rapid rate up their respective nerves. In addition some increase will take place in the number of resonators which are performing vibrations above the threshold amount. With further increases of intensity the number of resonators which are being effectively stimulated will increase while those which have previously been responding will send more frequent all-or-none impulses up their respective nerves. These processes will continue as the intensity of the tone increases until the nerves corresponding to the "in-tune" resonator are carrying the most frequent number of impulses of which they are capable. When this point is reached further increases of intensity will be perceived by a spreading of this "saturation" effect to the groups of resonators on either side, so that the group gradually grows with the corresponding increase in size of the effectively stimulated area as a whole. So far as theoretical grounds are concerned, the limit would be reached when the whole of the cochlea from end to end was receiving effective stimulation. According to Guernsey (15), 100 gradations of intensity, on the average, can be detected at any particular tone near the middle of the musical range. On theoretical grounds, the number might be a great many more than this.

REFERENCES

1. Baginsky, B. *Sitzber. Akad. Wiss., Wien,* 1863, 685.
2. ————. Die Function der Gehörschencke. *Arch. f. Anat. u. Physiol.,* 1883, **94**, 65-85.
3. ————. *Centbl. f. Physiol.,* 1888, 299.
4. Banister, H. Auditory theory: a criticism of Professor Boring's hypothesis. *Amer. J. Psychol.,* 1927, **38**, 436-440.
5. Barton, E. H. Textbook on sound. London: Macmillan, 1922. Pp. 687.
6. Boring, E. G. Auditory theory with special reference to intensity, volume, and localization. *Amer. J. Psychol.,* 1926, **37**, 157.

7. Burnett, C. H. The ear: its anatomy, physiology, and diseases. Philadelphia, 1877. Pp. xix+615.
8. Cosens, C. R. G., & Hartridge, H. A vindication of the resonance hypothesis of audition, IV. *Brit. J. Psychol.,* 1922, **13**, 48-51.
9. Everett, J. D. On resultant tones. *Proc. Phys. Soc. Lond.,* 1895, **14**, 93-102.
10. Ewald, J. R. Zur Physiologie des Labyrinths. VI. Eine neue Hortheorie. *Pflüg. Arch. f. d. ges. Physiol.,* 1899, **76**, 147-188.
11. ————. Zur Physiologie des Labyrinths. *Pflüg. Arch. f. d. ges. Physiol.,* 1903, **93**, 485-500.
12. Fraser, —. Deaf mutism. *J. Laryng.,* 1922, 137.
13. Frey, H. Die physiologische Bedeutung der Hammer-Ambossverbindung. *Pflüg. Arch. f. d. ges. Physiol.,* 1911, **139**, 548-561.
14. Gray, —. *J. Laryng. & Otol.,* 1905, **20**.
15. Guernsey, M. A study of liminal sound intensities and the application of Weber's Law to tones of different pitch. *Amer. J. Psychol.,* 1922, **33**, 554-569.
16. Hardesty, I. On the nature of the tectorial membrane and its probable rôle in the anatomy of hearing. *Amer. J. Anat.,* 1908, **8**, 104-179.
17. Hartridge, H. A vindication of the resonance hypothesis of audition. *Brit. J. Psychol.,* 1921, **12**, 142-146.
18. ————. A vindication of the resonance hypothesis of audition, III. *Brit. J. Psychol.,* 1922, **12**, 362-382.
19. Helmholtz, H. L. F. On the sensations of tone. (Trans. by A. J. Ellis.) (3rd ed.) London: Longmans, 1895. Pp. xix+576.
20. Hurst, C. N. A new theory of hearing. *Trans. Liverpool Biol. Soc.,* 1895, **9**, 321-353.
21. Keith, A. Appendix to: Wrightson, T. An inquiry into the analytical mechanism of the internal ear. London: Macmillan, 1918. Pp. xi+254.
22. Knudsen, V. O. Sensibility of the ear to small differences of intensity and frequency. *Phys. Rev.,* 1923, **21**, 84-102.
23. Liebermann, P. V., & Révész, G. Die binaurale Tonmischung. *Zsch. f. Psychol.,* 1914, **69**, 234-255.
24. Luciani, L. Human physiology. (Trans. by F. A. Welby.) Vol. IV. London: Macmillan, 1911-1921.
25. Mach, E. Bemerkungen über die Function der Ohrmuschel. *Arch. f. Ohrenhk.,* 1874, **9**, 72.
26. Mayer, A. M. Researches in acoustics. *Amer. J. Sci.,* 1894, **47**, 1-28.
27. Meyer, M. An introduction to the mechanics of the inner ear. *Univ. Missouri Stud.* (*Science Ser.*), 1907, **2**, 1-140.
28. Minton, J. P. Some physical characteristics of the ear. *Proc. Nat. Acad. Sci.,* 1921, **7**, 221-225.
29. Politzer, A. Textbook of diseases of the ear and adjacent organs. (Trans. by O. Dodd). London, 1894. Pp. xi+739.
30. ————. *Arch. f. Ohrenhk.,* 1864, **1**, 710.
31. Pavlov, I. P. Conditioned reflexes. (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927. Pp. xv+430.
32. Rayleigh, (Lord). Theory of sound. Vol. II. London: Macmillan, 1896. Pp. xvi+504.
33. Riesz, R. R. Differential intensity sensitivity of the ear for pure tones. *Phys. Rev.,* 1928, **31**, 867-875.
34. Rutherford, —. Tone sensation. *Brit. Med. J.,* 1898, **2**, 353.
35. Schäfer, E. A. Text-book of physiology. (2 vols.) Edinburgh, London: Pentland, 1898-1900.

36. Shambaugh, —, & Knudsen, V. O. *Amer. Otol. Soc.*, 1923.

37. Stepanoff, E. M. Experimenteller Beitrag zur Frage über die Function der Schniecke. *Monatsschr. f. Ohrenhk.*, 1888, **22**, 85.

38. Thompson, S. P. On the function of the two ears on the perception of space. *Phil. Mag.*, 1882, **13**, 406-416.

39. Toynbee, —. Diseases of the ear. London, 1868.

40. Waller, A. D. A possible part played by the membrana basilaris in auditory excitation. *J. Physiol.*, 1891, **49**, 1.

41. Watt, H. J. The psychology of sound. London: Cambridge Univ. Press, 1917. Pp. vii+241.

42. Weinberg, M., & Allen, F. On the critical frequency of pulsation tones. *Phil. Mag.*, 1924, **47**, 50-62.

43. Wilkinson, G. Analysis of sound by resonance. *J. Laryng. & Otol.*, 1922.

44. ————. Is the question of analysis of sound in the cochlea or by "central analysis" in the brain still an open one? *Amer. J. Psychol.*, 1927, **38**, 257-265.

45. ————. *J. Laryng. & Otol.*, 1927, **42**, 365.

46. Wilkinson, G., & Gray, A. The mechanism of the cochlea; a restatement of the resonance theory of hearing. London: Macmillan, 1924. Pp. xx+253.

47. Wrightson, T. An inquiry into the analytical mechanism of the internal ear. London: Macmillan, 1918. Pp. xi+245.

48. Yoshii, U. Experimentelle Untersuchungen über die Schädigung des Gehörorgans durch Schalleinwirkung. *Zsch. f. Ohrenhk.*, 1909, **58**, 201-251.

CHAPTER 8

THE CHEMICAL SENSES

GEORGE H. PARKER
Harvard University

AND

W. J. CROZIER
Harvard University

I

Sensory receptors naturally open to excitation by direct action of chemical agents are found in their most diversified forms, so far as can be known at present, among the higher vertebrates. It has been customary to speak of the senses of smell, taste, and the common chemical sense—olfaction depending upon receptors located in the mucous membrane of the nasal cavities, taste upon specific groups of sensory cells usually confined to the oral cavity, the common chemical sense upon more diffusely arranged nerve-terminals in moist mucous surfaces. Parker (87) has pointed out, however, that the "sense of smell" is probably served by differentiated receptors in the same manner as the "sense of taste" must be regarded as comprising perhaps four, and at the least three, different "senses." Connection with sensations is not a property of all kinds of receptors, and Parker has proposed the use of the term "recept" to denote the change occurring when the receptive branch of any reflex arc becomes active, without regard to connection with sensation. This is especially appropriate when it is remembered that, even in the case of such receptors as the eye, it is not feasible to consider that excitation and sensation are at all times synonymous. In the most general sense, recept was taken to include all effects resulting from reception of excitation to those changes in the central nervous system which include sensations when these are produced.

The discussion of any such topic as the present one has verbal difficulties as well as others. Terms it is necessary to employ are bound to imply one or another mode of viewing the relationships between excitation and sensation. We may as well try to remember that real acquaintance with sensation is confined to the individual, and a remoter kind of knowledge with respect to other human beings, with vaguer and increasingly more unsatisfactory guesses about other animals but that at the same time a completely "objective" terminology is impossible and a partial one rather grotesque as well as clumsy. The relationship between sensory excitation and sensation is rather simple, according to Adrian (2). The intensity of sensory disturbance is directly proportional to the frequency of impulses in the sensory nerve-fiber. By comparison with measured impulses in the nerves of cat and frog under comparable modes of sensory excitation, the curve of frequency of impulses with time of exposure is found to correspond to the requirement of the view that more intense sensation is correlated with increase and prolongation of sensory discharge. This is notably so for degrees of pressure excitation, and in the experiments on retinal excitation. Adrian takes these correspondences to signify that the "mental correlate" of intensity of excitation copies closely the physical disturbances in the sensory nerves, specifically their frequency. An effective *frequency* implies discontinuity, but an essentially steady state can be produced by mechanisms already recognized in the central nervous system, such as the presumably synaptic organ producing more gradual onset and decay of efferent discharge than is led into it by a given train of afferent pulses. The final integration obtained by such mechanisms could result in a smoothly declining or increasing intensity of central nervous disturbance corresponding rather closely to the course of changes induced in a receptor surface by a single step-wise alteration of the environment serving as a means of excitation, and it is not especially difficult to see how it could be made to follow more complicated disturbances.

These considerations have the merit that they utilize available quantitative information. They do not yet permit comparison of intensity of visual sensations, for example, with those of smell. The facts they are intended to summarize go only a little way toward making clear the basis of "sensation," but they avoid the speculative error of trying to determine, as it were *in vacuo,* the nature of sensation without real data concerning its arousal and its properties. It is not inconceivable that the differentiations among receptor cells of the chemical senses may involve and depend upon differences such as are revealed among receptors open to mechanical excitation. These are most conspicuously seen in the apparently characteristic relative rates of adaptation or sensory exhaustion under continuous stimulation, muscle spindle exhausting less rapidly than pressure organs, and these again less rapidly than tactile receptors or hair-base organs (2). The comparison is not too remote, because mechanical excitation must at bottom be dependent upon induced chemical changes, such as are brought about by reactive components when a solution acts upon a taste-cell or an olfactory cell.

From this general standpoint a sensation, as of taste, is the result of change of activity in a particular region of the central nervous system, the central terminus of a particular reflex arc. The classification of sense-organs in terms of the sensations they mediate is therefore bound to be deficient, since there is no ground for the opinion that the nature of a sensation is determined by the receptor organs (save in a remote and to a degree an accidental manner). The central connections of a given receptor are anatomically definite, but since not all receptors are connected with sensation recepts, such classification would fail doubly. The grouping of receptors is most effectively made upon the basis of their respective homologous excitatory agencies. The cogency of this procedure is quite apparent when it is attempted to deal also with the receptors of animals in general. The fundamental property of a receptor is its nervous connection with an effector part such as a muscle, gland, nematocyst, or other part the activity of which is dependent upon transmitted impulses. This is the zoölogically primitive property, and the only one generally possessed by sense-organs.

It has been maintained at various times that the elementary condition of a receptor is one in which a certain diversity of agencies may prove excitatory. The comparison with a protozoan such as *Amoeba* is enticing. But in fact there is good evidence of receptor differentiation even in sea-anemones, as Parker (86) has shown. In these animals, receptors in the outer epithelium are rather directly connected with muscles, so that in the main the nervous system approximates the diffuse elementary state from which it is possible to suppose that more complex nervous systems have arisen. Even here, however, the chemical excitation of the tentacles of the anemone by meat juice is not paralleled by any similar excitability of the pedal margin of the column; nor is tactile sensitivity of the pedal margin at all reflected to any similar degree in the responses of the tentacles. It is true that with sponges (83) no nervous elements at all can be recognized, so that the contractile cells must be regarded as directly excited to activity. But there is no clear evidence that the earliest, or at least simplest, receptors presumed to have arisen as triggers connected to such contractile elements were other than already in a measure specialized.

It is true that so far as investigation has gone the general properties of protoplasm seem as evident in sensory cells as in, say, *Amoeba*. None the less, there is apparent in the case of such "generalized" sensory surfaces as the skin of echinoderms and mollusks (e.g., *Holothuria,* 15; *Chiton,* 4), a functional separation of receptors activated by mechanical means and by light or by chemical agents. While this evidence is derived chiefly from experiments aiming at differential exhaustion, and is thus imperfect, theoretically, it is sufficient to encourage the idea that a basis for sensitivity must be sought which has a specific character. We can imagine that mechanical features of structure may bring it about that a given receptor is more easily excited under common circumstances by mechanical distortion than by change of temperature, for example. The facts of heterologous stimulation require the assumption that the changes immediately responsible for setting in train a series of one or more impulses are determined by intrinsic "metabolic" processes. But the specificities evident in stimulation by light, or in the categories of taste-buds in man, equally require the presence of specific receptive substances or arrangements of materials upon which a specific alteration in the potential factor of some form of energy in the environment can easily produce a definite effect.

There are certain advantages in considering the term "stimulus," so sadly abused in many ways, as applying not to the environing disturbance impinging upon a receptor but rather to the primary change wrought by this disturbance within the receptor. "Excitation" might then be taken strictly as having reference to the changes resulting in release of impulses from the primary receptor to the paths of transmission. "Excitatory agent" could then be employed for the external disturbance. This general view has certain advantages in the treatment of sensory adaptation, "after-effects," and the like, but it need not be insisted upon.

Of the three general types of receptors conveniently recognized on the basis of the nature of the homologous excitatory agent, namely, *radioreceptors, chemoreceptors,* and *mechanicoreceptors,* the chemoreceptors illustrate very clearly certain points already referred to regarding sensory differentiation. It is not to be supposed, however, that any fundamental distinction is to be drawn between the intrinsic activities of these organs. Mechanical stimulation must be regarded as dependent on chemical changes consequent upon the distortion or break-down of protoplasmic surfaces, and photic and thermal excitation upon the differential acceleration of processes in receptor cells absorbing the respective energies. Structural features which make certain receptors accessible to the action of external solutes, together with the possession of suitable reactive substances, are certainly concerned in the differentiation.

Among the chemoreceptors of vertebrates, three sorts of histological organization are apparent. They may be taken to represent stages in evolutionary change and elaboration (85). The neurons of the olfactory epithelium (and of the vomeronasal organ) are "primary sense-cells"; the perikaryon is located in the epithelium, their transmitting axons passing as nerve-fibers to the central nervous system. The receptors of the common chemical sense are "free" nerve-terminations in the epithelium of moist surfaces at apertures of the body, without clearly specialized end-organs; the perikarya are deeply situated, centrally from the epithelium. The receptive cells of taste-buds are "secondary sense-cells," forming groups of elements associated with the terminals of neurons of which the cell-bodies or perikarya are, as with the afferent neurons of the common chemical sense, deeply situated in association with the central nervous system.

Because of its similarity with the receptor element in such forms as sea-anemones and other invertebrates, in that its nerve-cell body is located in the receptive epithelium, the olfactory receptor is regarded as primitive, morphologically. From it, by centripetal migration of the perikaryon, the histological form of the free nerve-terminations of the common chemical sense may be considered to be derived. The formation of taste-buds apparently represents the result of a process termed by Parker (85, 87) "sensory appropriation." It is as if the presence of free nerve-terminations of the appropriate sort resulted in the determinate modification of ordinary epithelial cells, such that certain of them assume the form of receptor cells, the "secondary sense-cells" already referred to. The nature of this process is clearly evident when gustatory nerves supplying taste-buds are cut; the degeneration of the peripheral, isolated portion of the nerve is followed and accompanied by the degeneration of the taste-bud cells and the elimination of their remnants (74, 75, 76, 77); the regenerative growth of the cut fibers brings about the re-formation of taste-buds with the arrival of the terminals of the fibers at the germinative layer of the epithelium.

These phenomena raise a number of problems as to the nature of the metabolic and morphogenetic significance of the sensory neuron and its several parts. Parker (88) has re-examined the general question of the rôle of nerve-cell body and neurofibrils, and brings forward the probability that with regard to nervous operations these are likely to be solely of a nutritive function. Along such channels as the neurofibrils may provide it is possible to conceive that movement of substances may occur centrifugally, resulting in a kind of secretion from the ends of the terminal divisions of the neuron at the epithelial surface. Such substances may be regarded as responsible for the morphogenetic changes which result in the formation of taste-bud cells from ordinary epithelial cells. Moreover, if we consider the nucleus and perikaryon as a focus of origin for substances (oxidative catalysts, for example) of special importance in metabolism, it is possible to understand the comparative sensitivities of the three large categories of chemoreceptors. It will be pointed out that chemoreception apparently involves reaction between excitatory solute and constituents of the receptors. If these constituents are supplied through the ordinary progress of the native metabolic processes of the nerve-cell bodies—and we must suppose that all such substances are constantly being supplied and simultaneously removed, in the case of any active cell-constituent (79)—then it is reasonable to find that, broadly speaking, the olfactory receptor is the most sensitive, the taste-bud less so, the receptors of the common chemical sense least sensitive, even when the test substance is the same with the three sensory fields (90). In the taste-bud it is the taste-cell, rather than the primary neuron, which presumably supplies the essential reactants.

The three categories of chemoreceptors have it in common, among the vertebrates, that they are excited by substances in solution. With taste-buds this is quite obvious. Both in fishes and in air-inhabiting vertebrates this is also true of the olfactory organs of the nasal cavities, for with the latter the olfactory terminals are at least for the greater part imbedded in a layer of mucus, an aqueous solution of glycoprotein. According to Hopkins (49) the fine stiff ends of the olfactory cells in the frog project into the air space of the nasal cavity, but even if this condition be in fact a general one among air-dwelling vertebrates the exciting substance must first diffuse into the surface of the olfactory cell. Direct contact with the appropriately reactive substance is necessary for the excitation of a chemoreceptor. In the case of the terminals of the common chemical sense, the immediately neighboring epithelial cells presumably suffer excitation, and through their activity (as by secretion of juices) secondarily excite the nerve-terminals.

The direct character of the action of the stimulating agent as a chemical reactant with the chemoreceptor would appear to offer an unusual opportunity for the investigation of the chemical mechanics of excitation. There has been a tendency to treat this question as perhaps best approached in terms of the electrical changes invariably accompanying excitation, with the implication often given that these changes are in some sense primary or fundamental. Electrical changes, however, must be due to differential concentration or accumulation of ions, and to the chemical processes yielding this result we must look for the initiation of excitation. Sense-organs should afford the most suitable places in which to investigate the mechanics and dynamics of irritability (17, 60), because we have in them sensibly "unifunctional" elements, with index-responses permitting experimentation in the undamaged organism.

II. Olfaction

The two nasal cavities of man are separated by a smooth median septum. Their lateral walls exhibit a series of approximately horizontal folds, usually three in each cavity. The lowermost of these folds or nasal conchae extends over most of the length of the nasal cavity (Figure 1); the two conchae above this inferior one are of decreasing size; there is a certain amount of variation in the number of these folds (99). The conchae project into the nasal chamber in such a way as to provide an inferior,

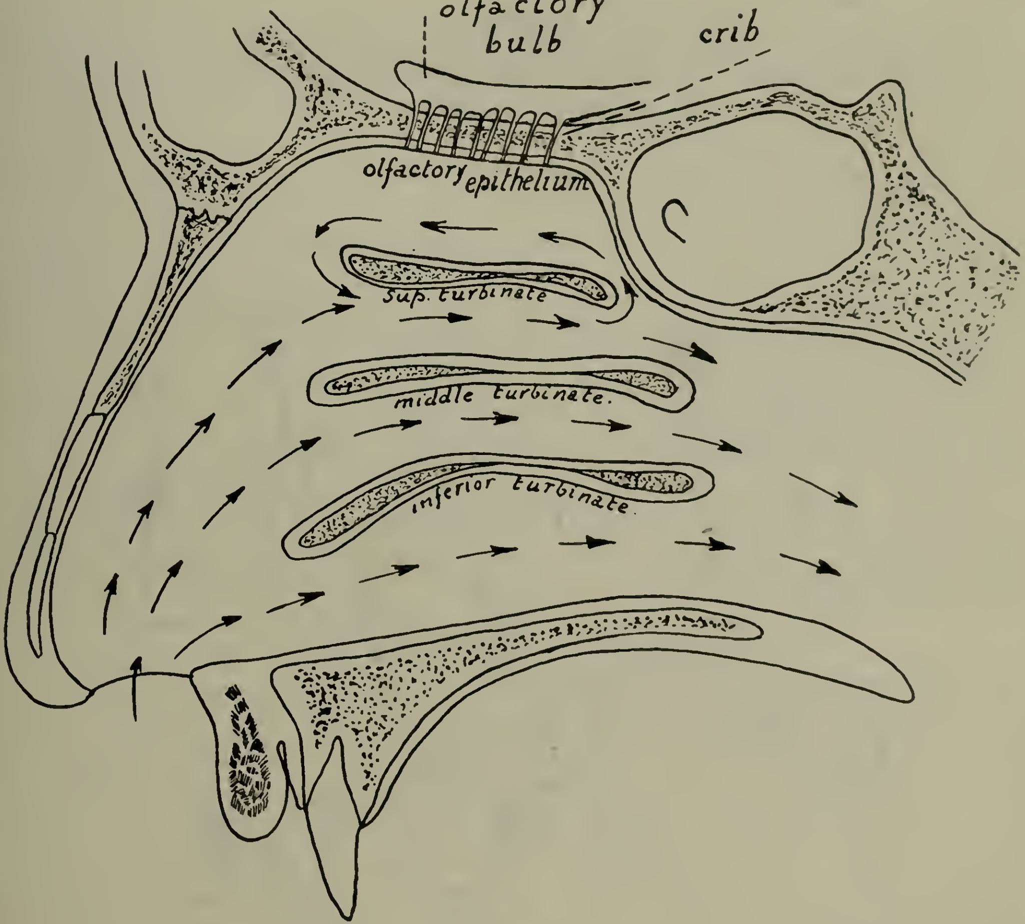


FIGURE 1

Anterior-Posterior Section through Nasal Fossae

The arrows indicate the direction of the air currents during inspiration. (After Howell.)

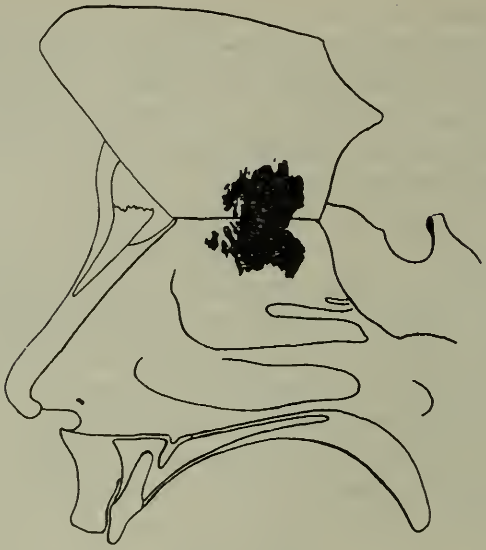
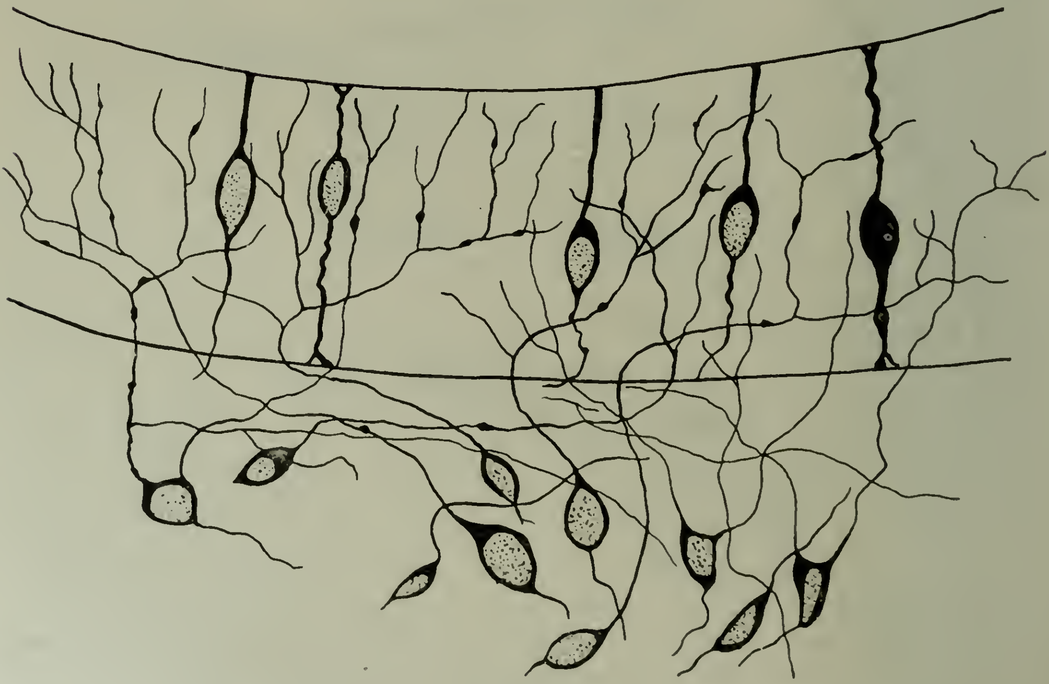


FIGURE 2
Olfactory Cleft of Man Opened by Turning the Nasal Septum Upward
The blackened area shows the distribution of the olfactory nerve.
(After A. v. Brünn.)



Olfactory Epithelium of an Embryo Chick (Ninth Day)
Showing olfactory cells, sustentacular cells, and free nerve-endings of fibers from ganglion cells of the trigeminal nerve. (After Rubaschkin, 1903, Figure 3.)

median, and superior meatus or channel. The inferior meatus first receives air drawn in through the external naris. Medially each meatus is confluent with a large common meatus, with which also communicates the so-called olfactory cleft lying between the most superior concha and the septum. These spaces are all rather freely connected with the choana (posterior naris) and so with the pharynx. A complex series of accessory spaces or sinuses of relatively large size, connected with the nasal cavities, are located in the bony framework of the nose.

The deeper part of the nasal cavity is chiefly covered by the so-called respiratory epithelium, which is pseudo-stratified, containing ciliated cells whose cilia beat toward the choana, and abundant branched alveolo-tubular glands containing serous as well as mucous cells. The conchae especially are highly vascular, their edges erectile and capable of enlargement through reflex excitation. Air entering the nasal cavity is warmed and moistened by passage over these surfaces, the rich circulation in which gives them a reddish hue.

The olfactory region proper is of a yellowish tint, or brownish. Only the inner face of the superior concha and of the adjacent outward face of the nasal septum is innervated by the olfactory nerve (12), forming the approximately rectangular olfactory area or von Brünn's area (Figure 2). The extent of this innervated surface is approximately 500 sq. mm., in man (92), but its boundaries are not sharply delimited. Ehrlich (28) and Arnstein (5) demonstrated that the "nucleated fibers" recognized by Schultze (1856) and other observers in the olfactory epithelium were continuous with nerve-fibers, and could thus be regarded as the olfactory sense-cells—a conclusion supported by later histological studies (95). These nerve-cells have their perikarya imbedded in an epithelium containing two or less numerous but distinct types of cells (Figure 3), the so-called basal cells next to the tapetum and the usual epithelial or sustentacular cells (the latter containing granules of yellowish-brown pigment distinguishing this region).

There is no reason to suppose that the free nerve-terminals also found in this epithelium differ essentially from those connected with the trigeminal nerve and found in other parts of the respiratory epithelium, or that they are significant for olfaction (98).

The cell-body of the olfactory cell is ovoid, containing a rounded nucleus with well-marked nucleolus. From the region of this enlargement, the cell extends distally in coarse rod-like form to a slightly swollen knob at the outer surface of the epithelium. From this enlargement, called by Van der Stricht (105) the "olfactory vesicle," spring the "olfactory hairs," cilia-like protoplasmic filaments (Schultze, 1862). The number of these "hairs" varies from 2 to 12 per cell, in different cases. Their delicate structure suggests that they are retracted and often lost in the course of making histological preparations, and Jagodowski (50) has indeed suggested (Figures 4 and 5) that they are more suitably termed flagellae or lashes, since in favorable preparations they may be two or three times as long as the body of the olfactory cell. Their importance lies in the fact

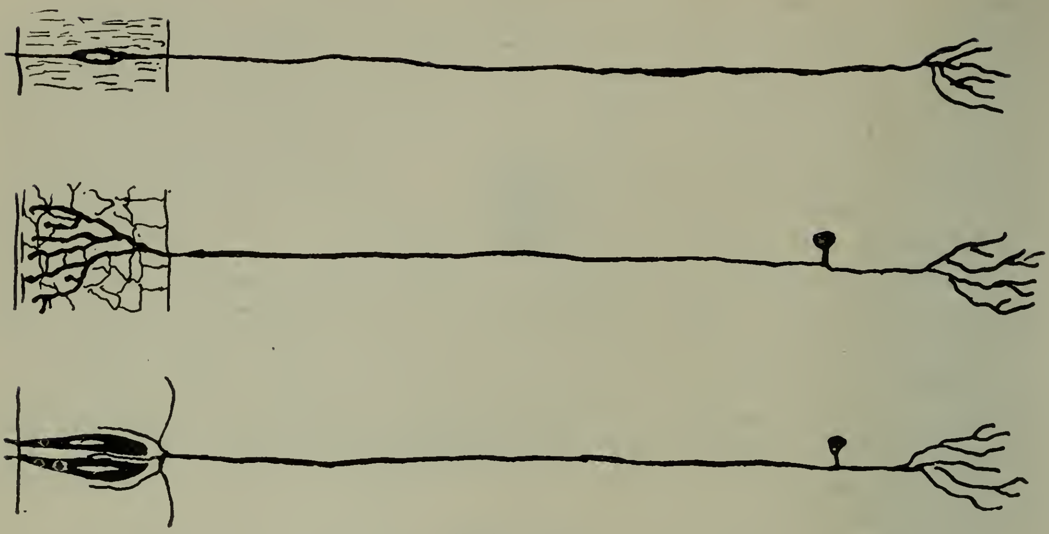


FIGURE 4

Schematic Representation of Typical Chemoreceptive Sensory Elements

Above, receptor of the olfactory epithelium and of Jacobson's organ; middle, receptor of the common chemical sense; below, compound receptor of a gustatory organ or taste-bud, comprising secondary sense-cells. (After Parker.)

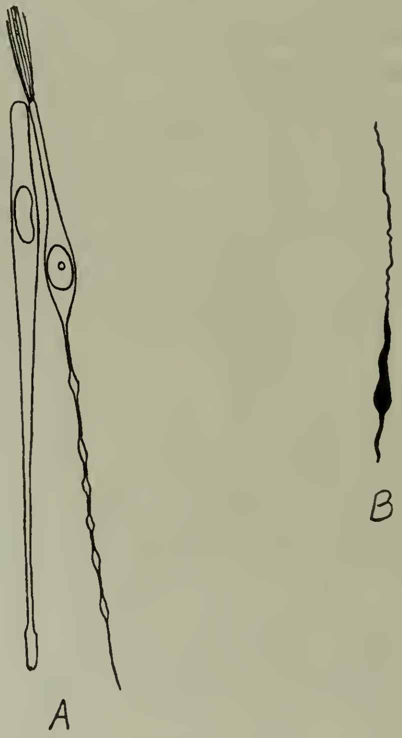


FIGURE 5

A. Isolated olfactory cell and sustentacular cell from a frog. (After Schultze 1862, Plate 1, Figure 4.)

B. A Golgi preparation of an olfactory cell from a pike (*Esox*), in which not only the cell-body and the basal nerve-fiber process are well shown but also the long peripheral olfactory flagellum. (After Jagodowski, 1901, Figure 8*c*.)

that they are undoubtedly the true olfactory surfaces, although imbedded in the mucus of the nasal cavity. The exact form of the olfactory cell is subject to a certain degree of variation (6), but the significance of this is uncertain.

In lower vertebrates less complication is evident in the form of the olfactory spaces. Keen-scented mammals, on the other hand, usually show an even more pronounced development of the conchae, which may be important for raising the temperature of sniffed air and thus assisting in the reaction of its odoriferous constituents with the olfactory epithelium. By contrast with macrosmatic forms, aquatic mammals (whales, porpoises) show practical degeneration of the olfactory sacs, at least in so far as concerns their function in smell. In bony fishes such as the catfish (*Amiurus*) the pair of olfactory pits are with rare exceptions on the dorsal surface of the snout, each pit being provided with an anterior and a posterior aperture (sometimes with tubular projections, as in the morays). Through these pits a current of water passes, as a result either of the activity of cilia within the pit or of the muscular action of pharyngeal muscles (as in such forms as *Fundulus*). Among elasmobranchs the nares are on the ventral surface, each pit being divided by a fold so arranged as to permit the passage of water through the olfactory cavity when the fish swims forward (Figure 6).

Even in elasmobranch fishes the nasal sac shows incipient division into two regions which have been taken to be prophetic of a division which is

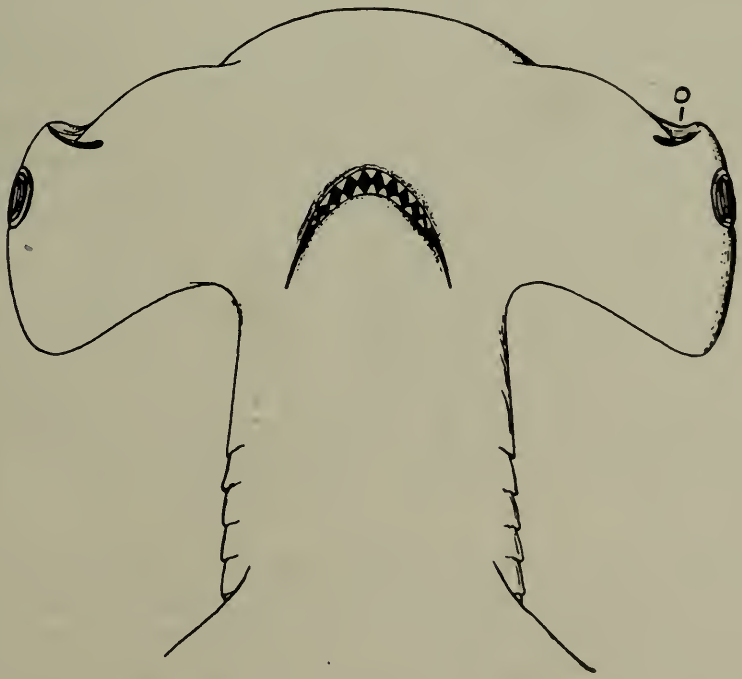


FIGURE 6
Ventral View of the Head of a Hammer-Head Shark (*Cestracion*) Showing the Olfactory Pits (*o*) Widely Separated
(After Garman, 1913, Plate I, Figure 2.)

more pronounced in frogs and in snakes. One member of this pair is the nasal cavity already discussed, the other, medially situated one, is the vomeronasal organ (Jacobson's organ). In snakes it opens independently into the oral cavity, a general arrangement which is retained in most mammals. In certain rodents, and in some primates (man) it opens by way of a proper duct into the nasal cavity. Von Brünn and later workers (92) showed that the epithelium of this sac is innervated just as is the olfactory epithelium. It has been suggested that the vomeronasal organ may be concerned with the sensing of conditions in the nasal secretions (11), but it seems usually to contain air, so is best regarded as a subsidiary olfactory receptor.

In still lower vertebrates, and particularly with invertebrates, the conditions of the olfactory receptors are quite different and diverse. Moreover, the means available for their identification have only begun to yield results which are of interest. This is notably the case with insects, for the behavior of which the significance of chemical excitation appears to be in some respects paramount. For certain insects at least, there is evidence (69, 33) to support the contention that smell and taste are as distinct as in higher vertebrates. Whether in such forms there is something corresponding to the common chemical sense (39, 40) is as yet quite uncertain.

Well-controlled experiments with hive bees have enabled von Frisch (32) to conclude that the organs responsible for associative responses to odor-boxes are in this insect located on the antennae. Removal of a portion of the antennae at a time showed that the essential region of these organs is confined to the fourth flagellar segment. The presence of but three-quarters of the extent of this segment suffices to permit good olfactory discrimination. The retention of the whole of this segment allowed the retention of an established odor-habit, and even the establishment of a new one through training. The receptors here concerned are the "pore-plates." These are ovoid plates of chitin surrounded by a thin narrow ring, under which the sensory fibers terminate. To these organs may be added the "olfactory cones" on the antennae, and certain "hairs," also regarded as organs of smell (32, 108). Among a series of muscid flies exhibiting different habits as regards their food and oviposition requirements, Liebermann (57) demonstrated a correlation between the number of antennal sensory cones on the antennae and the probable relative difficulty of finding adequate pabulum. Flower-feeders had, on the average, 1972 sense-organs; carrion- and excrement-feeders, 3,584; and bot-flies, 6,383.

On the other hand, McIndoo claims (64, 65) that the olfactory receptors of insects (bees, for instance) are not located on the antennae, but are found in the "olfactory pores" situated for the most part on the wings, at their bases, and on the proximal segments of the legs, as well as occurring scattered on other surfaces. The chief argument advanced in favor of this view, which has not found general favor, is obtained from measurements of reaction-time. According to McIndoo, in bees the speed of response to odors is inversely proportional to the total number of pores present in the several castes. There is no real reason to suppose, however, that reaction-time (latent period) for a response should depend directly on the number of receptors activated. McIndoo and others (67, 1) have shown, however, that in other insects olfaction is not confined to the antennal organs, although elimination of the antennae seriously reduces the reactivity to odors.

The certainty with which definite responses to odors can be obtained from certain insects, with proper control of experimental conditions, enabled von Frisch (31) to show that the sense of smell in these forms shows singular parallelisms with that of man. This was notably so in experiments designed to test by his training method the ability of bees to discriminate among odors such as those of essential oils. The judgment of similarity as to odor as made by man, and the delicacy of discrimination, turned out to be very much the same as evidenced by the behavior of trained bees. Moreover, although the details need not here detain us, von Frisch was able to show (for general summary see 34) that pairs of unlike substances smelling nearly alike to man are responded to very similarly by the bees in discrimination experiments. Such results tend to give promise of generality to any theory of specific olfactory excitation.

The movement of air through the nasal chambers of man in quiet respiration does not ordinarily excite olfaction, unless the concentration of the excitant be fairly high. Paulsen's (1882) tests with the split head of a cadaver showed that gentle currents drawn in through the nostrils did not impinge above the mid-region of the cavity, and thus tended to escape the olfactory area, although eddy currents may easily be set up. Sniffing movements, however, produce disturbance adequate to rise at most to the lower margin of the superior concha, even with an inspiratory movement (113). Eddy currents of this sort probably result in gradual diffusion of odoriferous material into the region of the olfactory cleft, at least in sufficient amount to excite more vigorous sniffing movements. The condensation of excitatory particles—it would appear first particularly in water vapor in the nasal cavity—probably serves to concentrate the stimulating agent on the olfactory mucous surface, and this condensation seems to be facilitated by the presence of such agents (113). Evidence of a reverse kind points to the necessity of solution of gaseous molecules as prerequisite for olfaction, since the time for the disappearance of olfactory sensation due to a series of odoriferous substances is directly proportional to their vapor tensions (111).

The quantities of substance requisite for olfaction are well known to be minute. Comparative measurements of odoriferous excitation have been made by means of the device arranged by Zwaardemaker, of which various modifications have been employed. The excitatory substance is spread upon the inner surface of a tube, into which fits a second, graduated tube supposedly free from excitatory substance. The extent of surface of the outer tube exposed by extracting the inner one (Figures 7 and 8) will be a measure of the substance given off for excitation, other things constant. The free end of the inner tube is adjusted to the nostril. Arbitrary units of surface used in this way are termed (Zwaardemaker) olfacties. The comparison of odors is possible, in an empirically semi-quantitative way, by the use of several such instruments, or by a double one permitting assault upon each of the two nasal chambers simultaneously. An instrument possessing theoretically more desirable features has been used by Allison and Katz (3) as an odiometer. Gas flow-meters are arranged to provide measured volumes of air passing in contact with a desired substance at a uniform rate, which can then be diluted with a measured flow of pure air, the mixture being delivered at the wide end of a funnel.

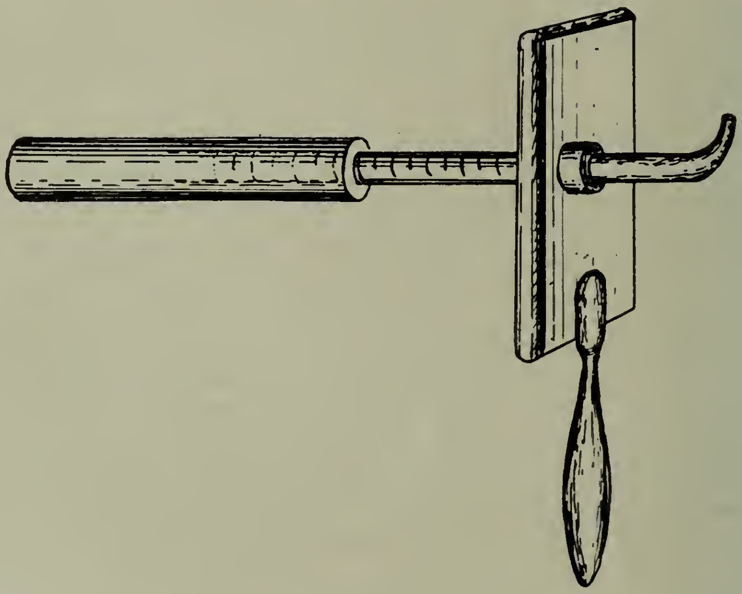


FIGURE 7
Simple Olfactometer
(After Zwaardemaker, 1895.)

Although the data resulting from attempts to measure absolute limits of olfactory acuity exhibit numerous discrepancies, depending in part upon technical conditions, they agree in pointing to the very small quantities of material concerned. When stated in terms of weight of substance inspired, the concentrations are of deceptively tenuous order; the numbers of

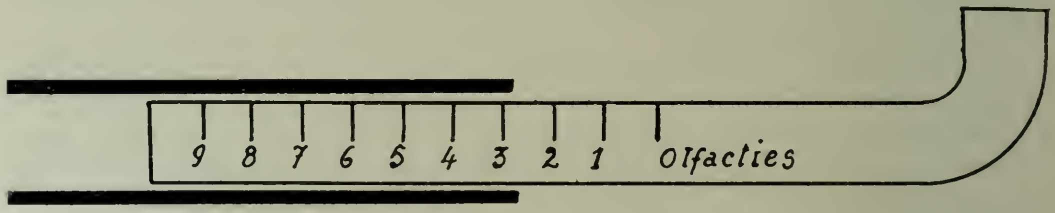


FIGURE 8
Simple Rubber Olfactometer
(After Zwaardemaker, 1895, Figure 14.)

molecules concerned are in fact large. We may recall that a quantity of energy equal to about one quantum per second suffices to excite the retina, for example. The experiments of Fischer and Penzoldt (29) showed that mercaptan could be recognized at a dilution of about 1/23,000,000,000 milligram per cc., or approximately 1/460,000,000 milligram for the nearly 50 cc. of air required for olfaction; according to von Frey, this minimum quantity of mercaptan amounts to 200,000,000,000 molecules. Certain minimal detectable concentrations of odorous materials are listed in Table 1.

TABLE 1

Concentrations of Odorous Materials to Give Detectable Excitation, Mg. per Liter of Air

(Allison and Katz, 1919)

Ethyl ether	5.83
Chloroform	3.30
Ethyl acetate	0.686
Ethyl mercaptan	0.046
Pyridine	0.032
Oil of peppermint	0.024
Iodoform	0.018
Methyl isothiocyanate	0.015
Butyric acid	0.009
Allyl isothiocyanate	0.008
Propyl mercaptan	0.006
Amyl thioether	0.001
Artificial musk	0.00004

In general, such experiments show that the higher members of homologous chemical series of compounds are more effective as excitants than the lower. The relative concentrations of the materials in a given series required for minimal excitation are retained for other arbitrary degrees of intensity of stimulation. According to Backman (7), the relative excitatory power is determined not alone by the aqueous (mucous) covering of the olfactory elements but also by the probably lipoidal character of the surface of the receptor. There is a certain qualitative justification for this view, since substances very little soluble in water are as a rule ineffective, but as in the case of the majority of substances which penetrate cells readily the chemically non-polar property which gives them solubility in non-polar liquids (lipoid solvents) seems more important (55, 54; for a recent bibliographic summary regarding permeability, 36). The solubility of olfactory excitants in water is of course a primary requisite in the case of aquatic vertebrates, such as fishes (84, 87). Even quite recently the opinion of Weber (109) is repeated, to the effect that gases or vapors are the homologous stimuli for smell, rather than solutions, for air-dwelling forms. While in a sense this notion may be correct as applied to insects, for example, it is certainly true that the olfactory organ of man can be stimulated by solutions (106) and presumably it is normally so excited as a rule.

In common with other receptors, continuous excitation of the olfactory organ induces sensory adaptation, such that a given odor is no longer perceived. In general, recovery of sensitivity takes longer than direct adaptation, but this must be seriously influenced by the movement of mucus charged with the stimulating substance away from the olfactory region. If during a period of continuous excitation of fixed intensity one determines the minimal intensity for stimulation, it is found (113) that the rate of sensory exhaustion increases with the adapting intensity. The

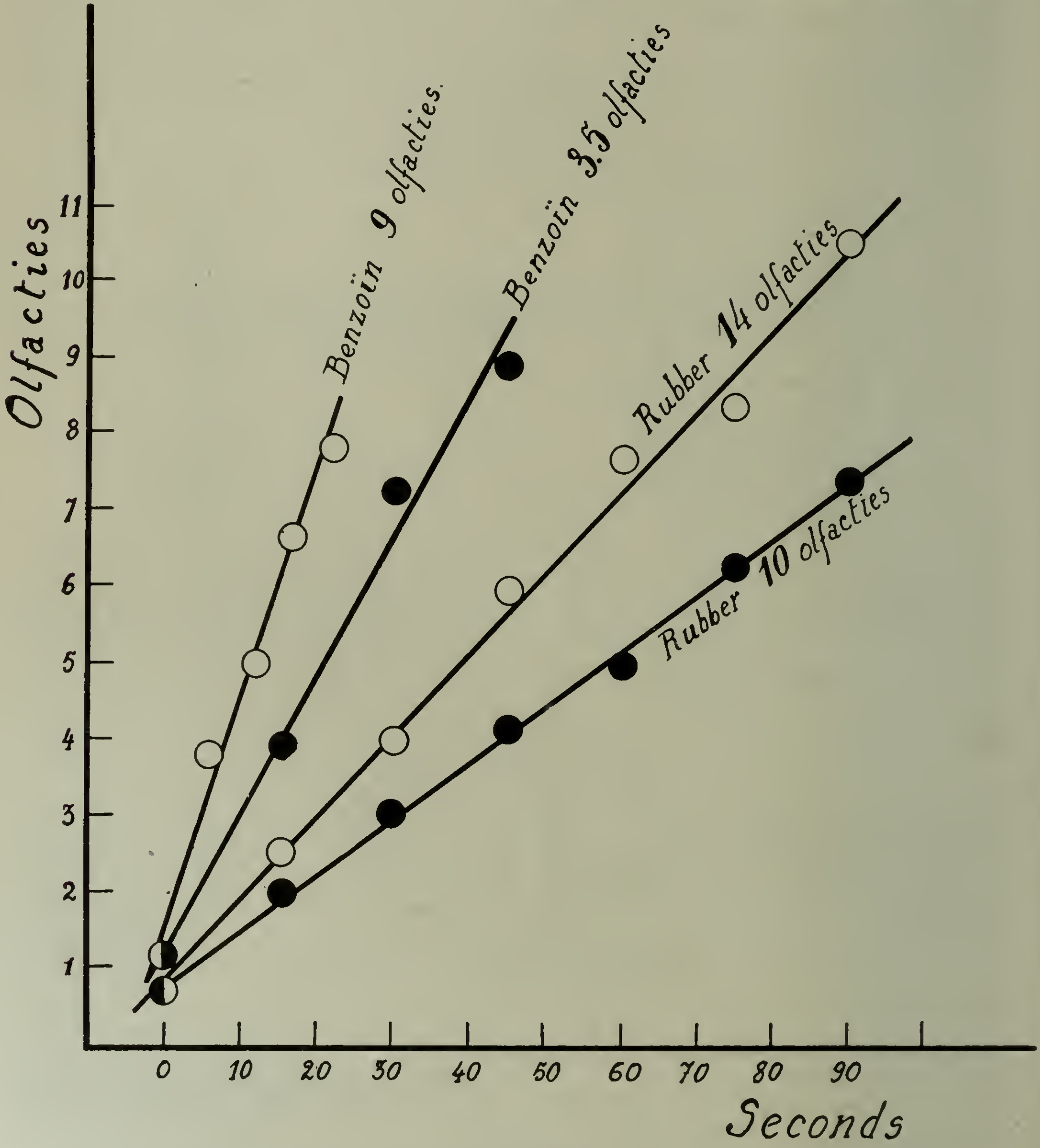


FIGURE 9

Curves of Olfactory Exhaustion Produced by the Action of Two Different Substances Acting for Different Periods

The threshold values (in olfacties) are marked on the ordinates, and the duration of stimulation (in seconds) on the abscissas. (Data from Zwaardemaker, 1895, Figure 22.)

time-course of adaptation, as determined by this method, is approximately rectilinear over the ranges which have been explored (Figure 9). This might be interpreted as due to the removal of "receptive" substance at a constant rate; or better—defining the sensitivity at any moment as $E_t = \frac{1}{C}$ where C is the amount of material requisite for excitation—we can say that the sensitivity declines hyperbolically during continuous exposure.

The classification of odors might be expected to lead to a certain understanding of the mechanism of their specificity. It cannot be held, however, that this end has been achieved. The classification due to Henning (1916) is in some respects the most satisfactory, though it inevitably leaves much to be desired. Perhaps part of the deficiency of any such scheme arises from the fact that the odors listed are not those ascertained for chemically pure substances. And in that there would be experimental difficulty not unlike the attempt to prepare any organic substance in pure state; the elimination of the last traces of impurity is in many cases impracticable, and these very traces usually exert detectable and confusing effects, analogous to the catalytic influences of impurities in organic reactions. Henning considers a scheme of six fundamental classes, which offers a means of picturing the relationship of "mixed" or intermediate odors (Figure 10).

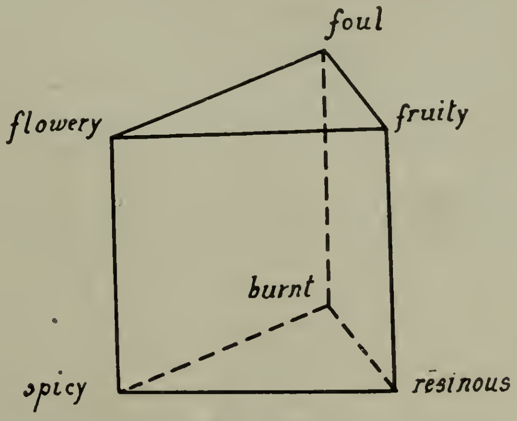


FIGURE 10
Henning's Olfactory Prism
(After Henning, 1916, Figure 4.)

Henning abandons the conception of particular osmophoric groups as determining the odors associated with particular molecules, and constructs his scheme for the aromatic compounds in terms of the structure of the ring molecule. As a rule the substituent groups which are odor-generating are the hydroxyl, aldehydic, ketonic, ester, nitro, and nitril, but the type of odor is determined chiefly by the total molecular configuration. These classes are labelled *spicy, flowery, fruity, resinous, burnt,* and *foul,* assumed interrelated as in Figure 10. The substances giving the spicy odor have, as in anisaldehyde, substituents in para-position; in the flowery group, in the meta- or ortho-positions (e.g., tuberon); the molecules of the fruity group have branched side-chains (e.g., citral); in the resinous, they are inner ring compounds (pinene); in the burnt group (as pyridine) the ring is simple; while for the foul odors (e.g., cacodyl) the ring is broken. This grouping is ingenious, though in part dependent upon conceptions of structure perhaps not always acceptable to the organic chemist, and it has the merit of emphasizing the reactive potentialities of the whole molecule of the exciting substance. But it tends to ignore the fact that excitation probably depends upon an inner mechanism, as regards the receptor, and that the "smell" of a substance is not primarily a property of the substance itself. Until more has been learned of the relationships between number of sense-cells, and their local distributions, excitable by typical groups of substances, progress here would seem likely to be slow; and the inaccessibility of the organ impedes such experiments.

As Henning recognizes, there is no real ground for assuming that his six classes correspond to elementary smells more than remotely analogous to elementary colors. A few cases of partial anosmia have been described, which suggests a functional differentiation of receptors such as might be responsible for the existence of integral olfactory qualities; but these cases are not numerous, nor particularly well examined. The existence of true olfactory components would not be negatived by the synthetic action of two or more mutually indifferent substances to produce a new odor, and such effects might receive ready interpretation in this way. The problem of competition of independently aroused odors, and of the mutual extinctions of smells received by the two nostrils simultaneously (113) is essentially one of central inhibition and lies outside the scope of the present summary. But the existing experimental facts do not justify speculation. Fusion-odors of this type have been observed, and their common existence in everyday life is a matter of universal experience. In this respect and in one other, olfaction contrasts sharply with its companion sense, taste. The greater delicacy of the olfactory receptor permits its function as a distance receptor, akin in a measure to sight, whereas with taste the source of stimulation must be close at hand, even in fishes, to provide a concentration of excitant adequate for stimulation. These two features of smell go far toward explaining the rôle which olfactory excitation occupies in the feeding and protective reflexes of higher animals, in the recognition of distant food and in the arousal of alimentary secretions, and in sex-recognition and excitement.

III. Common Chemical Sense

The general sensitivity of the skin of aquatic lower vertebrates to a variety of chemical irritants is well known (70, 81, 82, 101, 15). In man, this form of irritability is found on all exposed moist surfaces, including that of the nasal cavity. Fröhlich (35) showed that in the nasal cavity two sets of receptors were present which could be activated chemically—the olfactory receptors proper, and free nerve-terminals probably of the trigeminal nerve. For the latter, such irritants as ammonia are homologous stimuli, and they are responsible for more immediately visible reactions, such as sneezing and rubbing the eyes, than are seen with smell.

In the catfish *Amiurus,* Parker was able to demonstrate the persistence of responses to mildly irritating solutions applied to the surface of the body when the nerves connected with the facial and the lateral line nerves were cut, so that only the free nerve-terminals of afferent spinal nerves could be available for excitation (87).

There can scarcely be question of the distinctness of the human sensations attributed to the common chemical sense as contrasted with those of pain, smell, or taste. Functional separation of these receptors from those of pain is easily obtained by the use of cocaine as a differential narcotic. Similarly, differential exhaustion of an area of the skin of *Amphioxus* (81) by treatment with successive applications of weak nitric acid leaves the animal unresponsive to this form of excitation at that place, but fully responsive to touch. The existence of a separate series of receptors responsible for excitation by mildly irritating substances, the free nerve-terminals such as those of the trigeminal, seems therefore well established. The mode of their excitation differs from that of the two other chemical senses, inasmuch as it must be indirect, namely, through the activation of the general epithelium and secondarily of the nerve-terminals concerned.

IV. Taste

Aqueous solutions of a wide variety of substances, in moderately high concentration, excite the sensation of taste in man when applied to the tip of the tongue and to certain adjacent buccal surfaces such as the upper margin of the gullet, the epiglottis and soft palate, and certain less clearly defined regions of the upper pharynx. In fishes, Herrick (46) found many forms to possess taste-organs on the outer surface of the body, and in the catfish *Amiurus* these have been shown to be connected with the seventh nerve, so that from the standpoint of homology these must be regarded as gustatory organs (Figure 11). In *Amiurus* the discharge of meat juice in contact with the lateral surface of the body or with a barbel results in

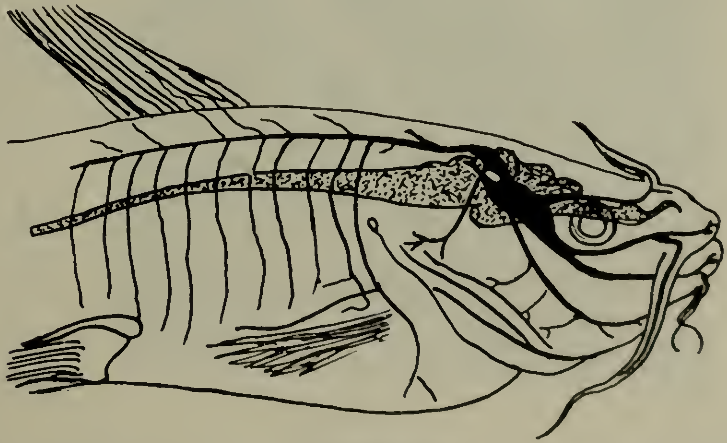


FIGURE 11

Lateral View of the Catfish (*Amiurus melas*)

Showing in black the gustatory branches of the facial nerve. (After Herrick, 1903, Figure 3.)

typical gustatory (feeding) responses (46, 85), with the implication of tactile excitation.

The taste-buds are spindle-shaped or bottle-shaped groups of cells often opening to the outside by a small pore (Figure 12). In the rabbit (41) compound buds with five or six pores have been found (Figure 13). In certain cases the outer pore leads through a short canal to the taste-bud proper. These buds are, in vertebrates, located on all surfaces from the

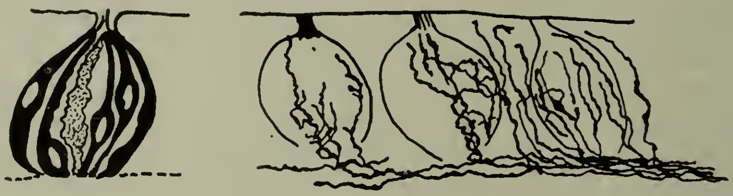


FIGURE 12

Sections through Taste-Buds of the Rabbit, Prepared by Golgi Method
Left (after v. Lenhossék), showing sensory cells and a "sustentive" element; right (after Retzius), showing terminal fibers.

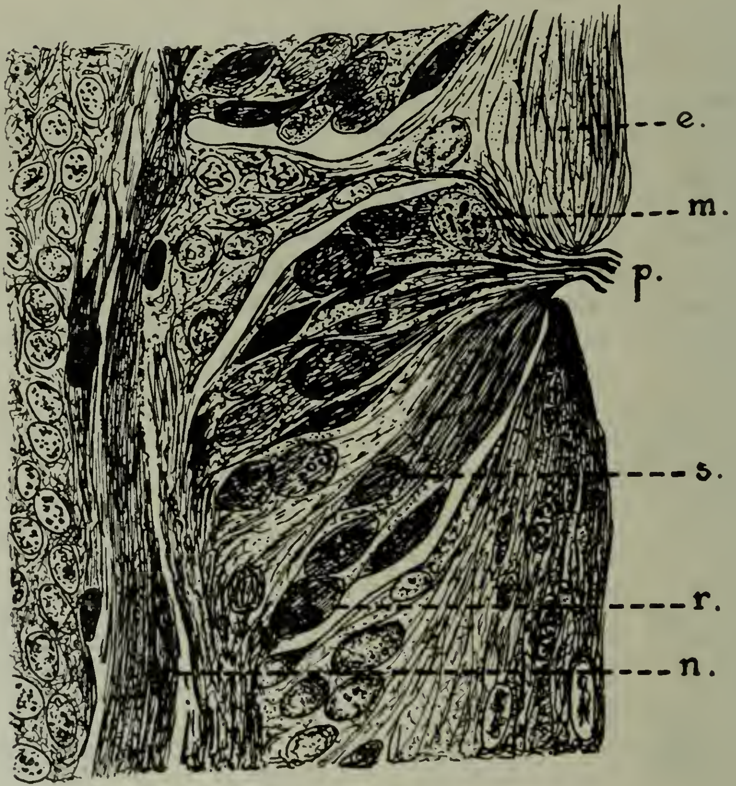


FIGURE 13

Section through One of the Taste-Buds of a Papilla Foliata of the Rabbit
e, superficial epithelial cells; *m*, leucocyte containing granules; *p*, gustatory pore; *s*, gustatory cell; *r*, sustentacular cell; *n*, nerve-fibers. (From Quain, after Ranvier, copied in Howell.)

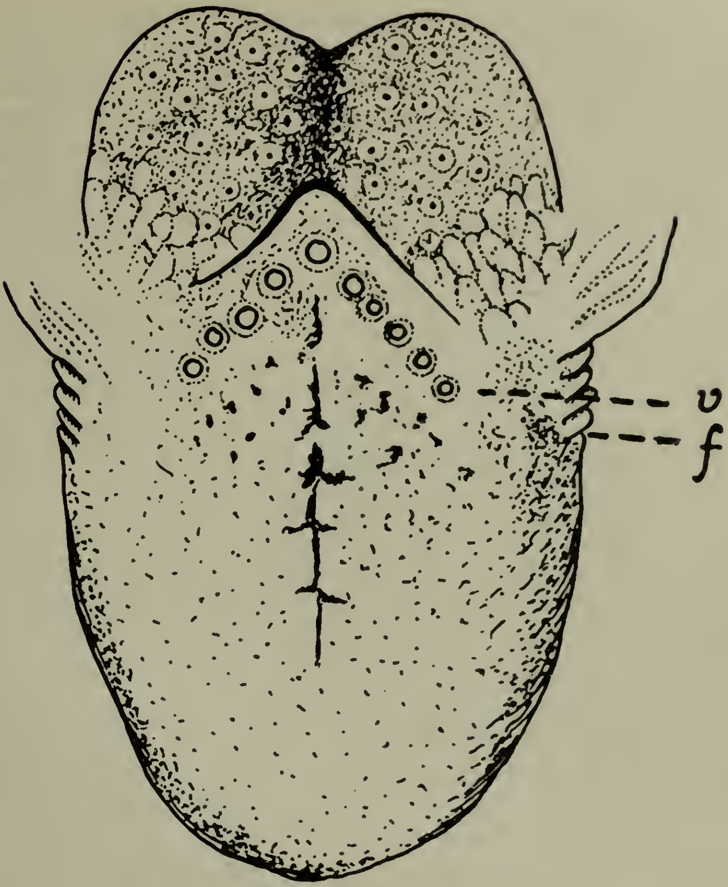


FIGURE 14
Dorsal View of the Human Tongue Showing Foliate Papillae (*f*) and Vallate Papillae (*v*)
(After Parker.)

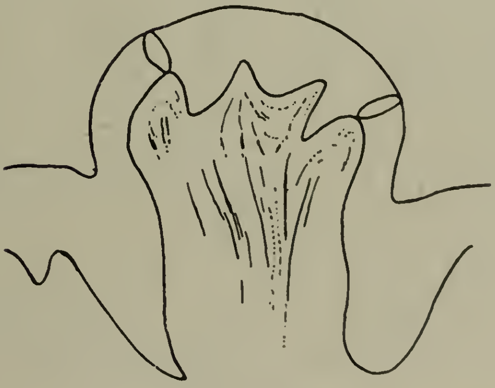


FIGURE 15
Vertical Section of a Fungiform Papilla Showing Two Taste-Buds
(After Parker.)

excitation of which gustatory responses can be obtained. The structure of the buds was first studied in higher vertebrates by Lovén (59) and Schwalbe (100), who showed that the "taste-cells" number from two to a dozen or so in each bud. Exteriorly the receptive end of the taste-cell projects into the pore of the bud, forming a "taste-hair." In addition to the taste-cells proper, each bud includes a number of sustentative cells. It is possible that these gradually replace the taste-cells functional at any one time.

Although in young children taste-buds are found on various oral surfaces from which they gradually disappear during growth, in the adult they are chiefly developed on the tongue, where they are almost exclusively associated with papillae of one or another form (Figure 14). Fungiform papillae on the dorsal surface of the tongue usually possess several taste-buds, located on smaller papillae dotting the upper surface (Figure 15). The foliate papillae at the lateral edges of the tongue are a set of from three to eight folds at the base of this organ. These folds are richly provided, laterally, with taste-buds. The vallate papillae form a chevron-like row of about 10 large but low elevations, each with a deep depression about it, in which the taste-buds lie (Figure 16). The most recent enumerations of these buds (42) show the number of these buds to be about 250 to each papilla but with much variation.

It was demonstrated by Retzius (94) that the nerve-fibers supplying the taste-buds are not directly united with the taste-cells, but end upon their surfaces or in close connection therewith, a view which subsequent work has confirmed. From the network of nerve-fibers immediately beneath a taste-bud, fibers of two sorts may pass to it—namely, those which end upon the surface of the bud and those which penetrate between its cells; the latter may be intimately associated with the so-called "supporting cells," and this fact is again consonant with the notion of a progressive replacement of generations of active receptors and with the state of affairs revealed in the degeneration of taste-buds separated from the trophic centers of their innervating fibers. This type of structure is practically uniform throughout the vertebrates.

Fibers innervating taste-buds travel in the vagus, glossopharyngeal. facial, and trigeminal nerves of man. The exact pathways of distribution of the relevant fibers are not yet entirely clear. Some from the vagus go to the posterior region of the tongue; the buds of the foliate and vallate papillae are supplied from the glossopharyngeal. The anterior region of the tongue is supplied by the combination of a branch of the trigeminal (the lingual nerve) and the corda tympani, but according to Cushing (27) the sense of taste persists on the anterior region of the tongue after removal of the nerve supply through the trigeminal (Figure 17).

A relatively small number of distinct taste-qualities are excitable by homologous stimuli for taste. The so-called "alkaline taste" was regarded by Oehrwall (71) as a complex excitation of one or more tastes with touch, and according to von Frey (30) weak alkalies such as excite this taste sensation (due to OH ion, 51) probably also produce olfactory stimulation, owing to their reactions with substances on the surface of the tongue. A "metallic taste" was recognized by Wundt (112) and others for the action of very dilute solutions of salts of heavy metals, but its nature is obscure and it presumably represents a complex of excitations. Mechanical stimuli are relatively ineffective, as are thermal changes also, in exciting taste-buds. The electrical excitation of taste-buds presents a number of complex features, in part dependent upon the accumulation of products of

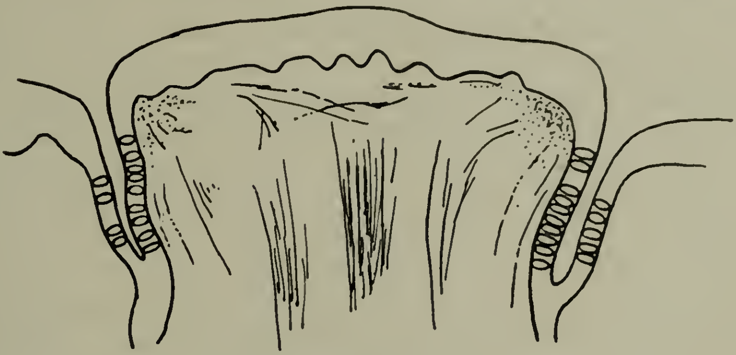


FIGURE 16
Vertical Section of a Vallate Papilla Showing Taste-Buds
(After Parker.)

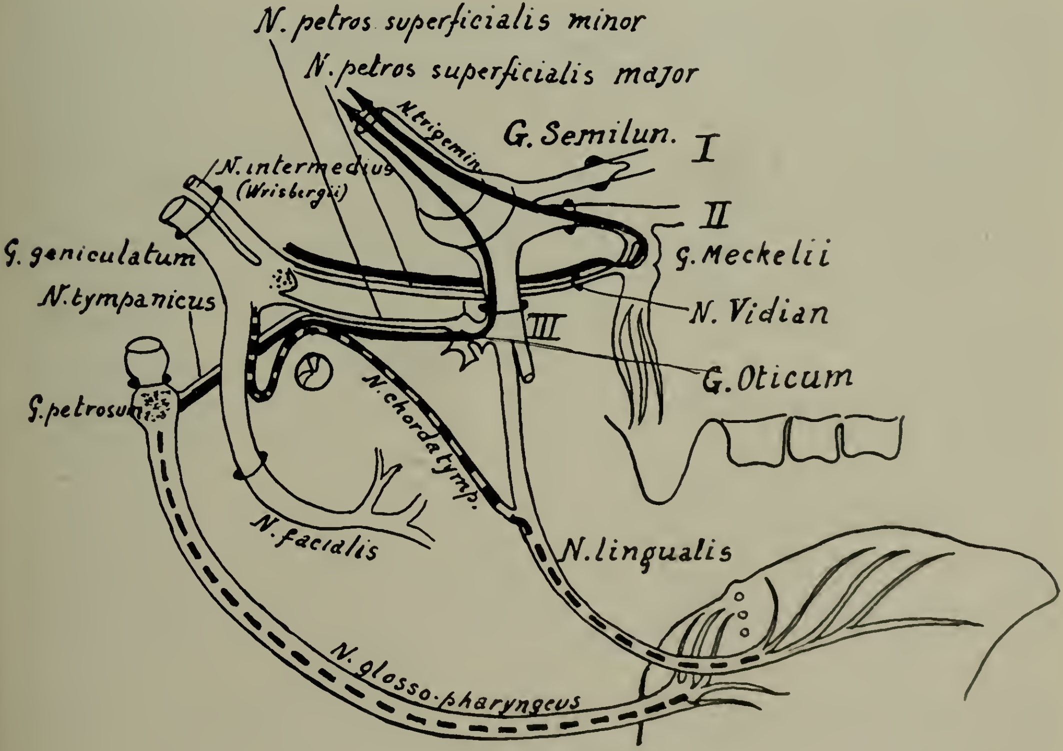


FIGURE 17
Schema Showing the Course of Taste-Fibers from the Tongue to the Brain After Cushing. The dotted lines correspond to Cushing's view; the plain black lines correspond to some authors' supposition. (After Howell, Figure 120.)

electrolysis and polarization, although this obstacle is not so apparent when rapid alternating current is employed. What would appear to be good evidence of direct electrical stimulation of taste-buds has been supplied by Parker and van Heusen (89) in the case of *Amiurus*.

The tastes easily recognized as distinct, namely, the sour, saline, bitter, and sweet, exhibit a certain homogeneity in the chemical character of the respective homologous stimuli for each. Henning (43) and earlier observers succeeded in showing that these four tastes correspond to local areas of distribution of taste-buds: the sour taste is represented on the lateral margins of the tongue, the saline at the tip and sides of the tongue, the sweet sharply maximum at the tip, the bitter in the region of the vallate papillae at the root of the tongue. The well-known work of Oehrwall (71) and of Kiesow (52) showed that individual papillae can be found in which responses to sour, saline, bitter, or sweet substances are alone recognizable. These results demonstrate the specific sensory qualities probably to be associated with definite taste-buds, a conclusion substantiated by the differential suppression of different taste-qualities when drugs such as gymnemic acid, stovaine, and cocaine are used. While it is undoubtedly true (Henning, 1916) that the tastes aroused by substances in any one group, such as the "sweet" or "salt," are not always alike, there has been a strong tendency (72) to regard these facts as signifying the presence of four definite gustatory senses. The comparative chemistry of taste excitation shows that regardless of this differentiation of tastes in man, animals in general are stimulated by sapid substances at minimal concentrations which are very similar (Table 2). In further support of this view are the latent periods measured for taste excitations (53), which in general show the latent periods to be greatest for bitter, next smaller for acids, next for sugar, and least for NaCl solutions. It is to be remembered, however, that until more extensive measurements are made with series of concentrations of representative substances, the weight to be

TABLE 2

TABLE INDICATING THE MINIMAL STIMULATING CONCENTRATION OF VARIOUS SUBSTANCES, FOR DIFFERENT ORGANISMS

(From Arey and Crozier, 1919)

Organism	HCl	NaOH or KOH	NaCl	KCl	Quinine	Picric acid	Author
Homo (mouth)	N/1,000	N/400	N/50	.....	M/25,000	....	Parker, 1912
Amiurus	N/20	N/100	N/50	.....	M/150	....	Parker, 1912
Amphioxus	N/500	.....	.....	.....	....	M/1,250	Parker, 1908
Ascidia	N/625	N/100	.....	N/4	M/2,500	....	Hecht, 1918
Balanoglossus	N/500	N/400	.....	N/200	M/1,000	....	Crozier
Aeolosoma	N/3,000	N/1,500	N/80	N/80	....	....	Kribs, 1910
Chiton	N/500	N/500	.....	N/160	....	M/1,500	Crozier & Arey
Chromodoris	N/700	N/200	.....	N/10	....	M/10,000	Crozier & Arey
Synaptula	N/600	N/200	N/4	N/40	M/10,000	....	Olmsted, 1917
Holothuria	N/500	N/500	...	N/500	....	....	Crozier, 1915

placed upon measurements of latency time in such cases must remain somewhat obscure.

Early observations on the taste of solutions showed colloids to be quite inactive in this respect, and that gustatory excitation is at least predominantly an effect of crystalloids: aqueous solutions of a number of salts—typically chlorides, nitrates, and sulphates of the alkaline metals, ammonium chloride, magnesium chloride, hydrochlorides of substituted amins, and certain bromides and iodides. Kahlenberg (51) and others found that solutions of NaCl so dilute as to be just perceptibly saline were more dilute than a solution of Na acetate equally saline in taste, for example, and hence concluded that the saline taste must be due to the chlorine anion, which also confers a saline taste in equi-dilute solutions of potassium or lithium chlorides. By similar tests, the order of effectiveness in exciting "salt" taste-buds was shown to be

$$Cl > Br > I.$$

The examination of other salts has shown (45) that with NO_3^- and $SO_4^=$ salts the anion is again the effective agent, but, aside from being more or less in line with other "anion effects" as to order of relative magnitudes in stimulating efficiency, little further can at present be said as to the mechanism of excitation. No evidence of ionic antagonism has as yet been

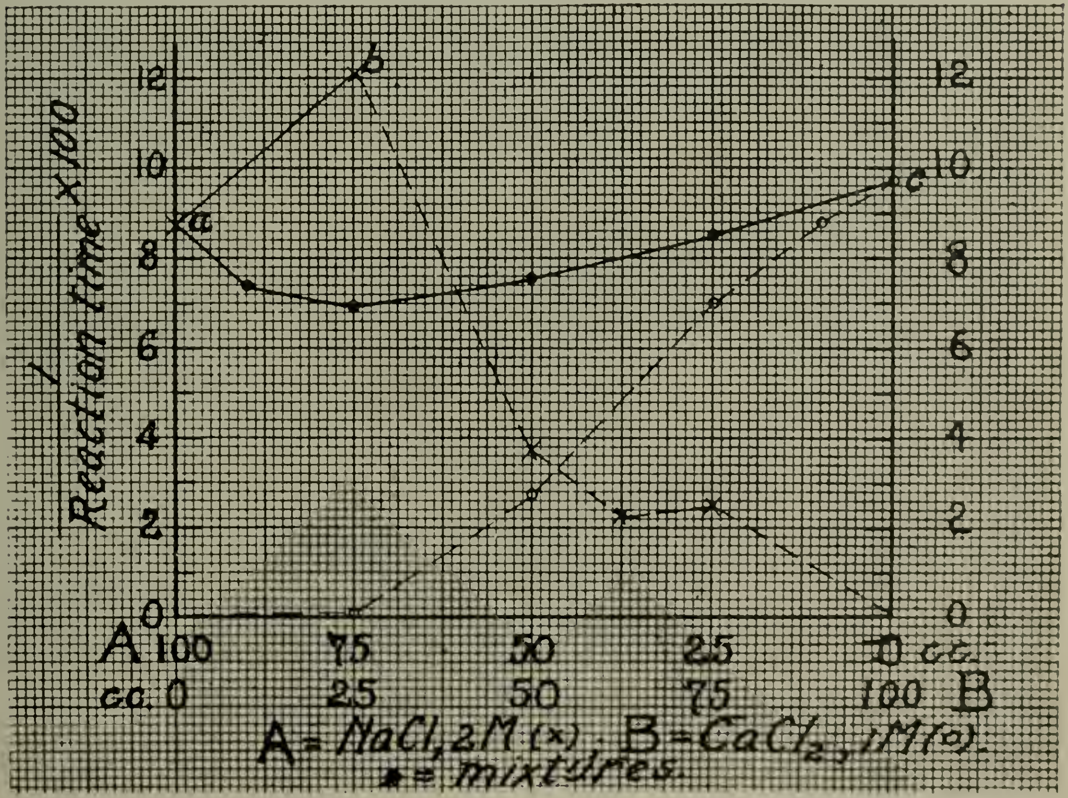


FIGURE 18*a*

REACTION-TIME OF THE FROG'S FOOT TO MIXTURES OF NaCl AND $CaCl_2$

definitely recognized in connection with the taste of salt solutions. In this connection it might be expected that investigations of the saline taste in solutions of different acidity might be illuminating. For the excitation of the common chemical sense-receptors of the skin of the frog's foot (17) typical phenomena of cation antagonism in solutions of chlorides have been measured (Figures 18*a*, 18*b*).

Somewhat more definite conclusions are possible for the "acid taste," as it may be called, for sourness seems exclusively excited by acid solutions (51, 97), and is thus to be attributed to the hydrogen ion. The hydrogen ion is in this respect effective in dilutions of acid at which the anion no longer excites the saline taste. Threshold concentrations for just-perceptible sourness have been determined for a number of acids. Table 3 is taken from a recent paper by Taylor (103). It has long been known that solutions of equal acid concentrations of different acids are not equally sour, and such facts as are represented in Table 3 show that the external H^+ concentration cannot be the solely significant variable (18). If it be assumed that, to excite a "sour" taste-bud, acid must combine with the surface of the taste-hairs, or even to penetrate somewhat to the interior of the taste-cell, further information should be obtainable from measurements

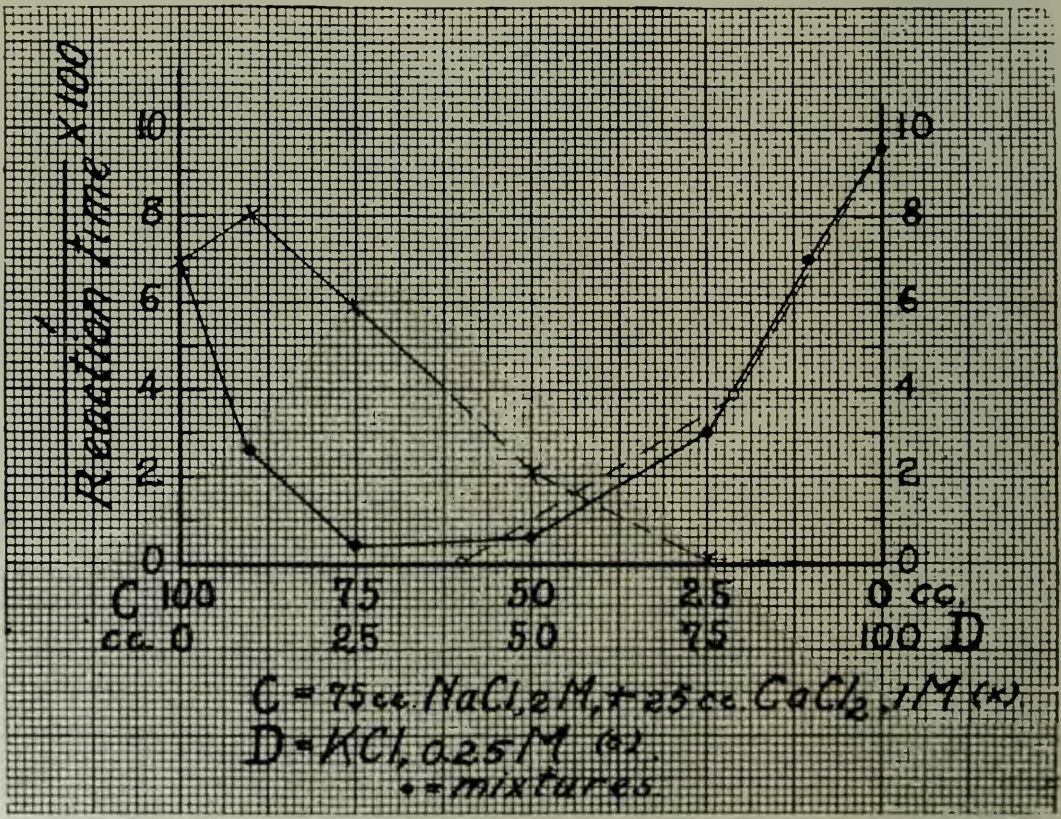


FIGURE 18*b*

Reaction-Time of the Frog's Foot to Mixtures of NaCl – $CaCl_2$ and KCl

TABLE 3

Threshold Concentrations for Sour Taste
(Taylor, 1928)

Acid	Concentration, *N*	H^+ ion concentration
Formic	0.0018	0.00055
Acetic	0.0028	0.00028
Butyric	0.0035	0.00027
Valeric	0.0037	0.00015
Oxalic	0.0020	0.00116
Succinic	0.0032	0.00034
Glutaric	0.0045	0.00034
Lactic	0.0028	0.00177
Tartaric	0.0022	0.00070

TABLE 4

Relative Concentration Gradients of Undissociated Acids Across Living Tissues to Produce Comparable Effects
(After Taylor, 1928)

Acid	Taste (Taylor)	Taste (Paul)	Excitation (Crozier)	Penetration (Crozier)
Formic	1.00	1.00	1.00	1.00
Acetic	0.17	0.43	0.28	0.60
Propionic		0.53	0.16	0.36
Butyric	0.17	0.068	0.16	0.24
iso-Valeric	0.21		0.14	0.10
Caprylic			0.075	
Caproic			0.006	
Bromoacetic		6.1		
Chloroacetic		7.5	4.17	5.95
Dichloroacetic			13.5	
Lactic	0.54	2.9		1.12
Benzoic				0.52
Salicylic				3.76
Succinic	0.34	0.48		0.93
Malic		1.9		4.26
Tartaric	1.60	1.1		8.00
Oxalic	6.3			6.5
Malonic				8.8
Glutaric	0.38			
Carbonic		0.019		

of the penetration of cells by acids. In certain cases this can be done with the aid of tissues containing natural indicator pigments (19-22, 25, 26), such as the epithelium of the mantle of *Chromodoris* (a nudibranch mollusk). Such data have only a remote connection with the problem of physiological permeability, although the point has several times been misunderstood by writers on that subject. If we assume that to give visible evidence of penetration a given acid solution must establish by diffusion an intracellular hydrogen-ion concentration corresponding to the color change of the indicator, computations can be made of the relative speeds of penetration as corrected for the intracellular titration of the indicator solution and thus of the relative ease of union with the surface of the cells

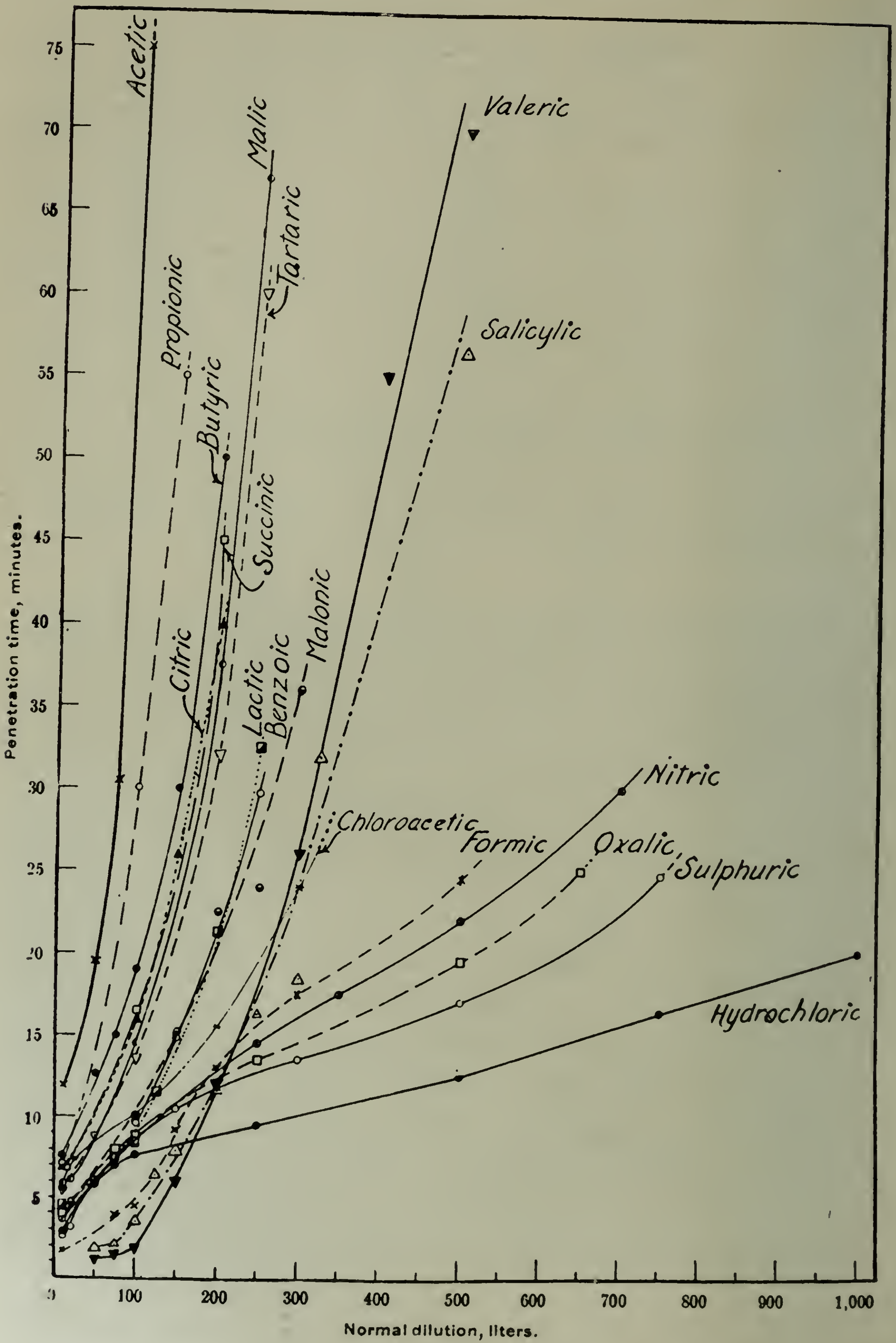


FIGURE 19

The Relation between Dilution and Time of Penetration (whether by chemical combination in the usual sense, or by simple solution). Another way of obtaining data from such measurements in a form useful for comparison has been carried out by Taylor, namely, by computing the ratio of undissociated acid concentrations existing external to the cells and internally at the time the indicator changes color (in this case, at about pH =5.6). These ratios may be assumed to represent the

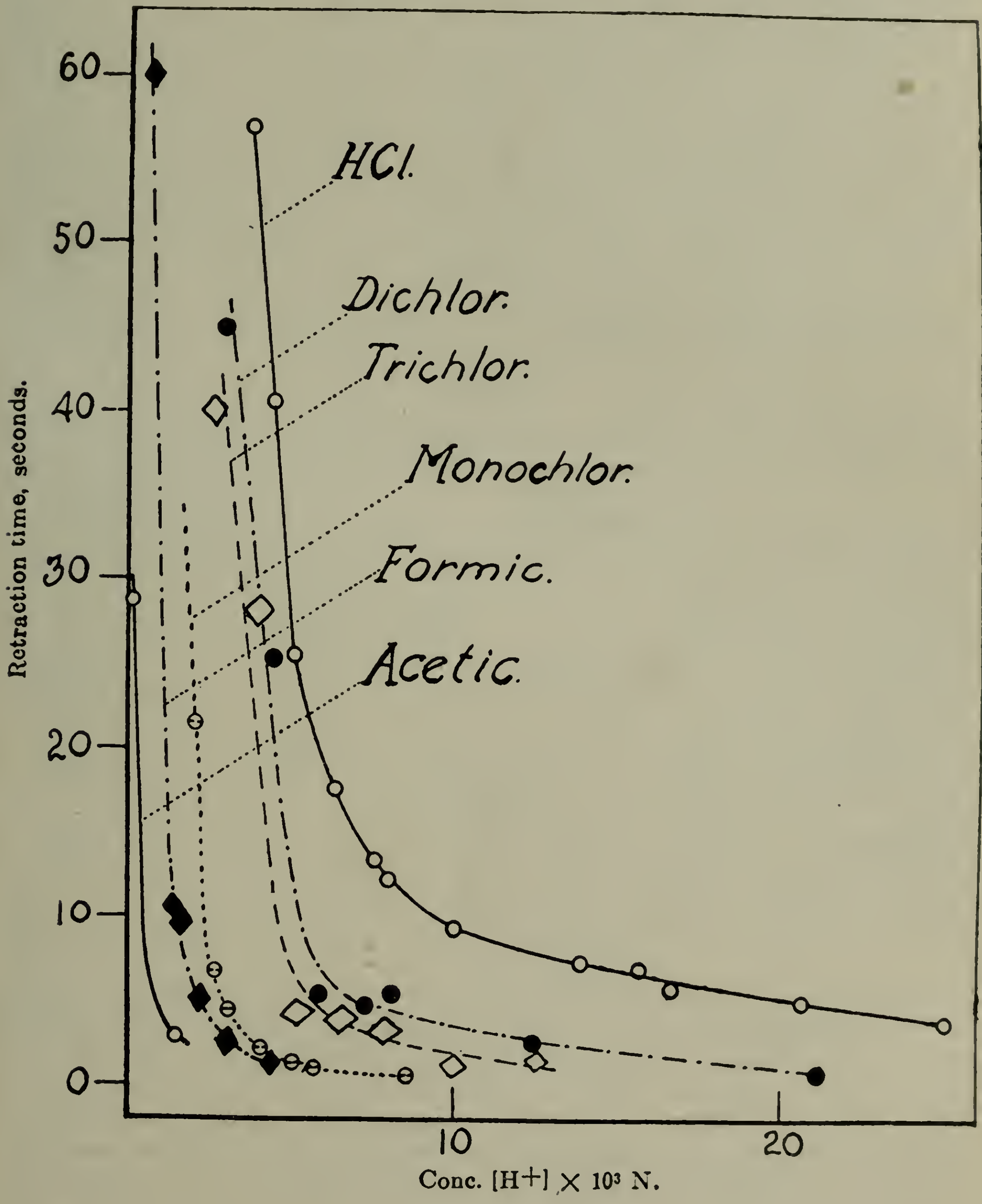


FIGURE 20
The Relation between "Retraction-Time" and the (Calculated) Hydrogen-Ion Concentration of Certain Acid Solutions

relative driving forces necessary to induce penetration to a fixed extent (i.e., producing equal change in internal pH). These figures are reproduced in Table 4, which also gives certain additional ratios computed for other series of acid-stimulation experiments (Figure 19).

These additional tests concern estimations of the strengths of acids required to produce equivalent sour excitations (91). The procedure is legitimate to the extent that if odor effects are eliminated the sour taste is the same with all acids, but it might not be true that measurements of "equally sour" concentrations should give the same relative order of concentrations at all (moderate) degrees of sourness. However, at least a crudely parallel case has already been noted in olfaction (cf. *supra*). There is also included the results of tests upon the stimulation of the manure worm *Allolobophora* (Figures 20 and 21), which again show parallel features of excitability. In this instance, the relation between speed of induced movement and the concentration of acid bathing the posterior surface of the body was measured with various acids (24). The pertinent results are collected in Table 5. They demonstrate the parallelism between velocity of penetration and sourness or exciting efficiency. With solutions of equal pH a high degree of sourness is associated with ease and speed of penetration. A notable instance is provided by CO_2. In considering such series of acids it is apparent that such highly polar groups as —OH and —Cl or —Br markedly decrease the ease of penetration, and within such a group as in the monobasic fatty acids there is close parallelism with "adsorption" effects known for these substances. The accumulation of non-polar groups in a lipoidal film, as at the surface of a taste-cell, should on Langmuir's conception lead to an almost linear relation between length of carbon chain and excitation power. The experiments with *Allolobophora*

TABLE 5

Acid	Concentrations of Acids (and H^+)			
	to produce equal sourness (Paul, 1922)		to give a retraction time of 10 secs. with *Allolobophora* (Crozier, 1918)	
Formic	$7 \text{x} 10^{-4}$ M;	$[H^+] = 3 \text{x} 10^{-4}$	$11 \text{x} 10^{-3}$ M	$[H^+] = 14.3 \text{x} 10^{-4}$
Acetic	22	1.9	32	7.6
Propionic	33	2.56	22.5	5.6
Butyric	5	0.96	23	5.7
iso-Valeric			18	5.3
Caproic			11	3.9
Caprylic			1	1.1
Monochloroacetic	12	8.0	8.4	29.1
Dichloroacetic			5.8	52.5
Monobromoacetic	11	7.2		
Lactic	23	5.0		
Succinic	8	2		
Malic	8	4		
Tartaric	4	3		
CO_2	53	0.4		

ora (23) permit semi-quantitative consideration of this point (Figure 22).

A further illustration of the rôle of combination with the receptor (penetration of its surface) is given by the behavior of alkalies. NH_3 (or NH_4OH) solutions are known to penetrate living cells much more quickly than does NaOH, much the stronger base; roughly, the ratio of speeds of penetration is about as 90:1. Tests with *Allolobophora* (23) show a direct proportionality between concentration and stimulating power for these two bases (Figures 23 and 24), their efficiencies being relatively as required on the theory of penetration.

The excitation of the sour taste may therefore be considered to depend upon the H ion, brought into the interior of the surface layer of the taste-cell by diffusion of the acid (either as undissociated molecule, 80, or as both ions it provides), and there reacting with some specific taste-cell mechanism. The speed with which the acid penetrates determines the intensity of the excitation.

The sweet and bitter tastes have in certain respects closely allied properties. Closely related compounds may be in one case sweet, in the other bitter, while with a number of sweet substances there is found (at appropriate concentrations) a bitter after-taste. Taylor (104) has suggested that a single receptor mechanism may be involved in the two cases, pointing out that Gymnema extract, for example, temporarily suppresses both

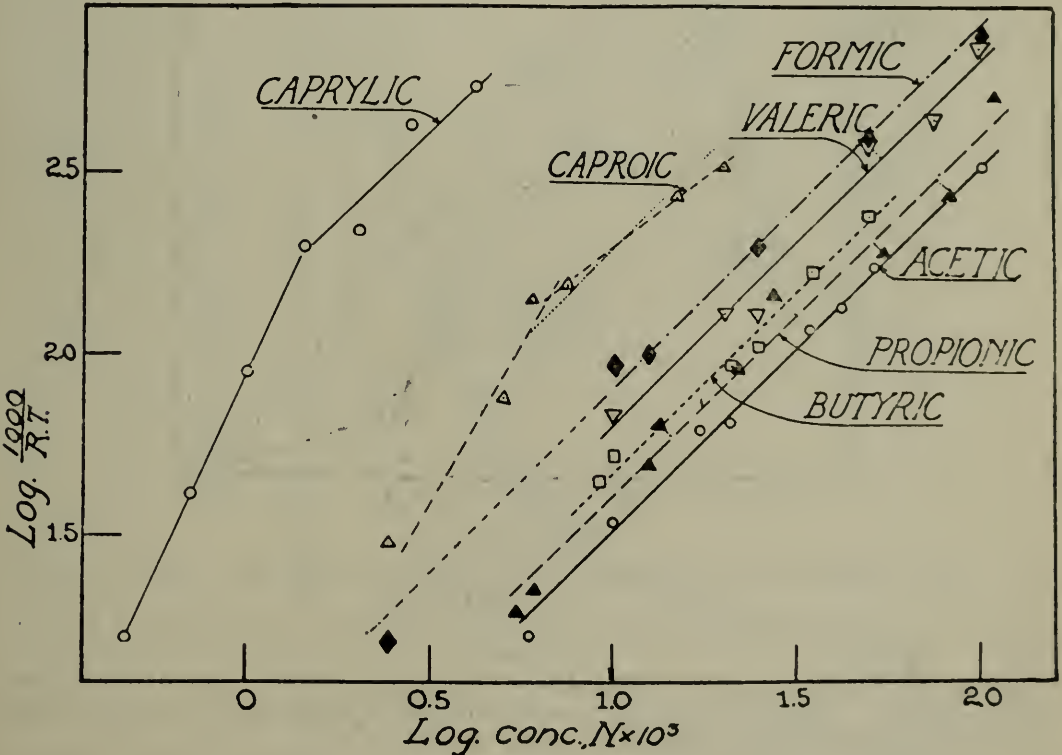


FIGURE 21

THE RELATION BETWEEN THE RECIPROCAL OF THE "RETRACTION-TIME" OF THE EARTHWORMS FROM SOLUTIONS OF FATTY ACIDS AND THE CONCENTRATION senses. Alkaloids and some glucosides, as well as bile salts, many plant extractives, and other substances typically exerting pronounced toxic effects upon cells in general are distinguished by a bitter taste. This is also true of a number of narcotics. These are usually supposed to have a special affinity for lipoidal constituents of the cell surface, and in general initially exert an excitatory influence. Taylor suggests that the bitter taste—more strictly the excitation of the "bitter" group of taste-cells—is due to an effect similar to that in the inhibitory action of the typical protoplasmic depressants, and that the sweet taste is, on the other hand, due to phenomena akin to those seen in the preliminary excitatory phase in the action of such substances. Anaesthetic or narcotic action has often been assumed due to absorption by cell lipoids. From the standpoint advocated by Taylor, the sweet taste should be characteristic of molecules very slightly soluble in

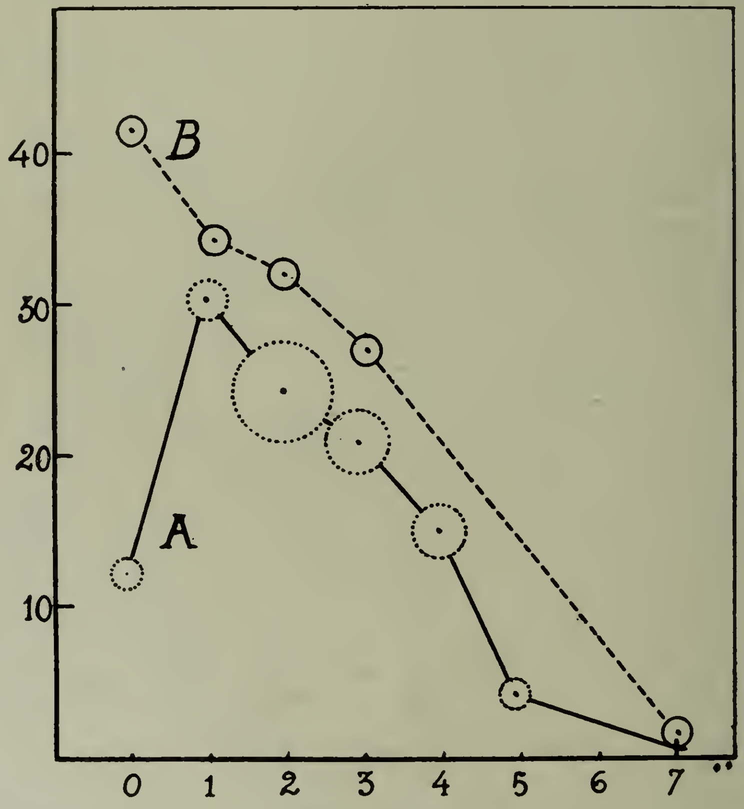


FIGURE 22

A. The ordinates are the concentrations (N x 10^3) required to effect a definite amount of stimulation (such that $\log \frac{1000}{R.T.} = 2.00$). The abscissas are the number of CH_2-groups in the molecules of the monobasic fatty acids from formic 0 to caprylic 7. The dotted circles are drawn with a radius equal to the average deviation calculated from individual observations.

B. Against the same abscissas as in *A* there are plotted the times required to effect membrane formation in 60 per cent of the eggs of the sea-urchin, according to experiments by Loeb, in acid solutions at 0.001 N concentration. The unit of ordinates is here 0.1 minute.

water. Very sweet substances such as the α-antialdoxime of perillaldehyde, roughly 3,000 times as powerful in evoking the sweet taste as sucrose, and n-hexyl chlormalonamide (about 300 times as sweet as a 6%-sugar solution), are very slightly water-soluble. Saccharine, with its intense sweetness, is again only slightly soluble. Among inorganic salts, magnesium sulphate is soluble in non-polar solvents and is bitter, while the closely related beryllium sulphate, not soluble in alcohol, for example, is sweet, like the salts of glucinum.

The substances exhibiting a sweet taste with bitter after-taste would on this view have their effects explained by assuming that in low intracellular concentrations they are less closely associated with surface lipoids, the taste becoming bitter when time permits their accumulation. Such materials are rhamnose, β-methyl glucoside, m-nitrobenzoic acid, several saccharine derivatives, and a number of others. There are certain inconvenient substances, however, such as ortho-benzyl-benzoate, of which the primary taste is bitter, the after-taste sweet. But in this general direction it may be possible to account for the fact that the separation be-

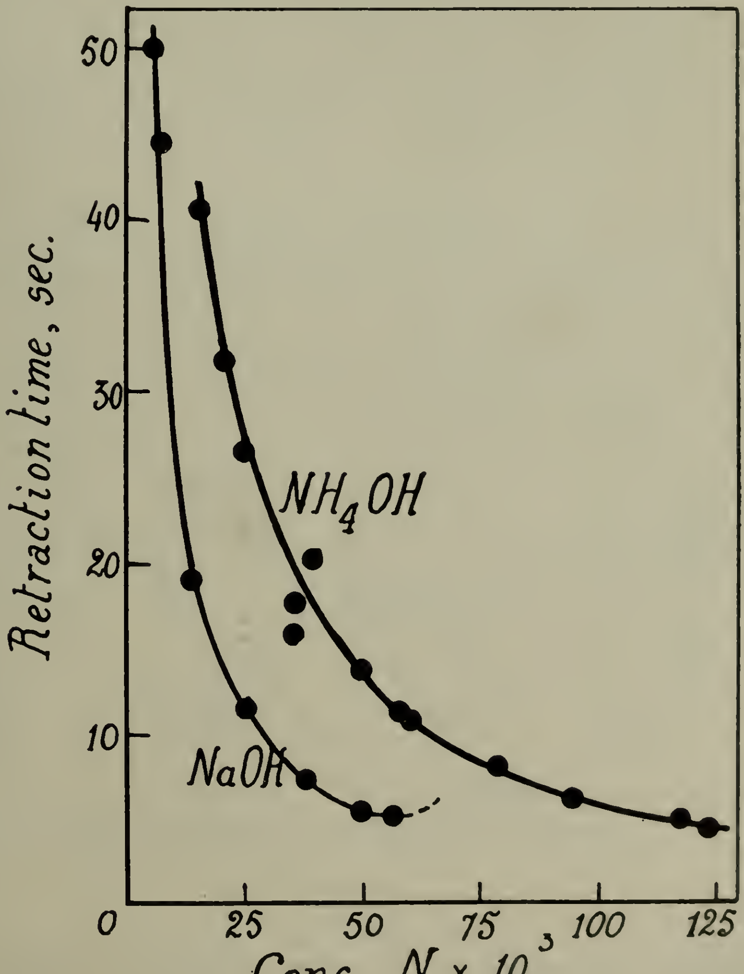
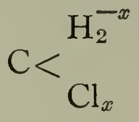


FIGURE 23
Retraction-Time of Earthworms from Alkaline Solutions

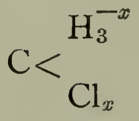
tween sweet and bitter tastes is not the same for all persons; to some even saccharine is said to be distinctly bitter. Variation among individuals in this respect might be interpreted in terms of different (perhaps temporary and fluctuating) states of structural organization of the taste-hairs as regards their lipoidal constituents.

Much information has been collected relative to the influence of substituent groups on the taste of organic molecules (8, 13). Oertly and Myers (73) have listed a number of sweet-producing and sweet-accentuating molecular arrangements and groupings (glucophores and auxoglucs). Radicals making for potention sweetness and containing polar groups are:

$—CHOH—CHO$
$—CHOH—CO=$
$COOH{\cdot}CHNH_2—$
$CH_2OH, \ CHOH—$
$CH_2ONO_2—$

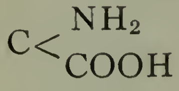


$$\mathrm{C}<\begin{matrix}\mathrm{H}_2^{-x}\\ \\ \mathrm{Cl}_x\end{matrix}$$



$$\mathrm{C}<\begin{matrix}\mathrm{H}_3^{-x}\\ \\ \mathrm{Cl}_x\end{matrix}$$

(Cl signifies any halogen)

Cohn (13) recognizes in this category also the group  $\mathrm{C}<\begin{matrix}\mathrm{NH}_2\\ \mathrm{COOH}\end{matrix}$

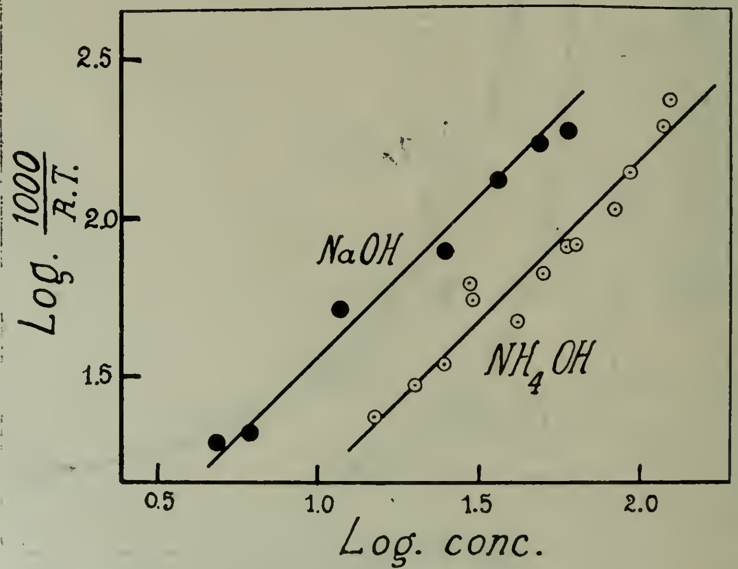


FIGURE 24

THE STIMULATING EFFICIENCY OF ALKALINE SOLUTIONS (MEASURED BY THE RECIPROCAL OF THE RETRACTION-TIME OF THE EARTHWORM), PLOTTED AGAINST CONCENTRATION OF ALKALI

The resolution of several substances exciting the sweet taste into glucophore and auxogluc moitiés, according to Oertly and Myers, is illustrative of their analysis:

Compound	Glucophore	Auxogluc
Glycol	$CH_2OH{-}CHOH$	H–
Glycerol	$CH_2OH{-}CHOH$	$CH_2OH{-}$
Fructose	–COCHOH– (H)	$C_nH_{2n+1}O_n$
Glycine	$-CHNH_2{-}COOH$	H–
Ethyl nitrate	$-CH_2ONO_2$	$CH_3{-}$

There are, however, at present too many puzzling irregularities in the taste-exciting properties, of organic molecules particularly, to permit useful rationalization of the details of gustatory excitation.

Reference has already been made to the fact that the stimulation of various invertebrates by direct application of fairly concentrated solutions to surfaces of the body shows a certain degree of uniformity (of course with minor differences) throughout the animal series in the way in which representative sapid substances act. Table 6 has been compiled by Köhler from recent results of various writers; the figures are, of course, subject to irregularity consequent upon the diverse methods of testing.

TABLE 6

LIMITING DILUTIONS FOR CONCENTRATIONS FOR STIMULATION (PERCENTAGE CONCENTRATION)

Animal	Sugar	Picric Acid (1-4, 9, 13*a*) Chinin (5-8, 10, 11, 13*b*)	H_2SO_4 (1-4, 9, 13*a*) HCl (86) Acetic Acid (5, 6, 8*a*, 10, 12, 13*b*) Tartaric Acid (11)
1. Asterias	7.5	0.0075	0.01
2. Fulgar		0.04	0.01
3. Limnea		0.05	0.05
4. Helix	8.0	0.15	0.065
5. Octopus		0.0005	0.006
6. Crangon		0.0005	0.01
7. Pyrameis	0.002		
8. Bee	2 to 4	0.04	*a*) 1.2 *b*) 0.045
9. Mustelis	50.6	0.38	0.065
10. Triton	2	0.012	0.06
11. Lacerta			0.51
12. Canary			0.5
13. Man	0.578	*a*) 0.0002 *b*) 0.00007	*a*) 0.006 *b*) 0.013

The situation among insects is of special interest, because of their feeding habits as well as of the large rôle played by responses to licked secretions in the establishment and maintenance of social habits (110). Minnich has employed forms with extensible probosces, individual hairs of which can by careful work be tested for gustatory function, as by the extension of the proboscis in response to contact with a sugar solution. By his training procedure with bees, von Frisch has been able to demonstrate taste sensitivity to sugars and discriminative ability between, for example, glucose, fructose, and lactose, galactose, and others. Minnich (68) has shown that with insects the taste-receptors are not localized upon the mouth parts, but are present also in the tarsal segments of the legs. In *Pyrameis* the sensitivity of the tarsal receptors was shown to depend quantitatively upon the state of feeding as regards sugar. When saccharose was withheld, solutions as dilute as M/12,800 were in some cases successfully distinguished from distilled water.

A certain analogy may be allowed between the differentiated gustatory papillae and the diversity of retinal elements concerned in the perception of color, with the added complication that in man the tasting of sapid materials involves the possibility of a shift of the material to different parts of the tongue. From the properties of the four, or if the sweet and bitter receptors be counted as one type fundamentally, then of three groups of receptors the properties of complex and fluctuating tastes must then be derived. Nothing is known as to the possible rôle of number of papillae excited in this connection. That the mechanism of excitation is fundamentally chemical there is little reason to question, but the part taken by the structural character of the taste-hairs as possibly modifying this chemical union is as yet rather obscure.

Accepting the view that chemical action of dissolved crystalloid particles upon receptive substances in the taste-cells is the basis of gustatory excitation, an attempt has been made by Lasareff (56) to give mathematical formulation to the theory of the process of this excitation. Such formulations represent necessary steps in the progress of interpretation, and as such they are to be welcomed. Too often the appearance of mathematical statements in the form of equations arouses antagonism and distrust. If it is remembered that the real intent of such statements is to increase clarity of understanding as a means of making further experimental analysis intelligent and fruitful, a large part of this difficulty may be avoided.

Lasareff assumes that the amount of substance in a gustatory element will determine the amount of decomposition products resulting from it when the taste-cell is brought into contact with an exciting solution. Let $[C]$ represent this concentration, $[C_1^1]$ the concentration of the decomposition products resulting therefrom; K_1 and K_2 will stand for the velocity constants of the decomposition reaction and of the elimination of the products, respectively. Assuming that the reaction with the exciting substance is a first order reaction (often termed, but without entire justification, "monomolecular"), then

$$\frac{dC_1}{dt} = K_1 [C] [C_2] - k_2[C_1^1],$$

where $[C_2]$ stands for the concentration of the stimulating solute. In other words, the velocity of accumulation of the reaction product will depend at any moment upon the difference between the speed at which it is being produced and that at which it is being removed from the local seat of olfaction (as by diffusion). If we let $[C_0] = [C] + [C_1]$ represent the maximum effective concentration of the receptive substance at any time t,

$$\frac{dC_1}{dt} + [C_1^1] [K_2 + K_1 [C_2]] - K_1 [C_2] [C_0] = 0;$$

the integration of this equation gives

$$[C_1^1] = [C_0] \frac{K_1 [C_2]}{K_1 [C_2] + K_2} \left\{ 1 - e^{-K_1C_2+K_2t} \right\}.$$

It may be assumed that to obtain a minimum of sensation with excitation a certain concentration of the exciting solute is required; let this be denoted by B. Then from the first equation given, the stationary state existing without excitation will be defined by an equality in rates at which the product of the receptive substance is produced and removed, and

$$K_1 [C_2] C = K_2 [C_1], \text{ or } [C_1^1] \frac{K_1C_2C}{K_2};$$

to effect excitation, the threshold defined by B must attain the value

$$B = \frac{K_1 [C_2] [C]}{K_2}.$$

The gustatory sensitivity may be defined as

$$E = 1/C_2$$

since obviously this sensitivity will be at least roughly in proportion to the dilution (i.e., reciprocal of concentration) of a given substance required to stimulate. Then

$$B = \frac{K_1C}{K_2E}, \text{ or } E = \frac{K_1C}{K_2B}.$$

These equations permit predictions of a certain kind concerning the processes of sensory adaptation in the case of the gustatory organs. Since this aspect of chemical excitation has been very little examined in a precise way, the indications thus obtained are suggestive. The decrease of $[C]$ under the action of a gustatory excitant corresponds, clearly, to a decrease in the measure of sensitivity, E. This is obviously the process of gustatory adaptation. On the present assumption, if E_0 be the initial sensitivity,

$$\frac{E}{E_0} = 1 - me^{-K_3 t}$$

should express the change in gustatory sensitivity (as measured by the effectiveness of a given exciting solute) during the progress of minuthesis or sensory exhaustion; here m is a constant the magnitude of which depends upon the degree of antecedent action of substances producing the sensation of taste, and must be regarded as including the effects, if any, of a central nervous character as well; K_3 is a collective velocity constant. Lasareff does not notice, however, that a formula of this sort is very commonly obeyed by a diversity of natural processes, and, while it follows from the assumptions used, its empirical validity would not necessarily justify them. Lasareff cites Table 7 as indicating the degree with which measurements of E, the gustatory sensitivity during continuous exposure to a substance such as sugar, show changes agreeing with those which may be calculated from this equation. The range of change is of course not wide, but the correspondence is interesting, although Lasareff's papers are singularly devoid of the information one would like to have concerning the way in which the measurements were actually obtained. In all probability, however, the course of gustatory adaptation is indeed an orderly one; and we need not be surprised if under the most carefully devised conditions, and with due allowance for the presence of epithelial secretions which may change in amount or in kind during stimulation of oral surfaces by solutions, the course of sensory adaptation is found to be similar to that in the retina (cf. Hecht, Chap. 5). In the latter case, only absorbable energy is delivered to the receptor organ, while with taste-cells a reactive component is permitted to bathe the sensory surfaces, so that with a photosensitive organ the experimental conditions are by comparison almost ideally simplified.

TABLE 7

Behavior of the measure of gustatory sensitivity, E, in terms of the equation $E/E_0 = 1 - me^{-K_3 t}$, where E_0 is the initial sensitivity. K_4 is an arbitrary constant.

t, secs.	0	90	120	180	210	250	360
K_4 $(1 - E/E_0)$, observed	100	95	90	89	81	75	68
K_4 $(1 - E/E_0)$, calculated	*100*	*91*	*89*	*84*	*81*	*78*	*70*

The notion of a receptive substance located in a taste-cell carries with it the idea that the qualitative differentiation of the taste-buds might be due to the fact that there are four separate receptive substances, appropriately segregated in distribution, corresponding to the four conspicuous gustatory qualities. This would require that substances evoking, let us say, the salt taste, or the sweet, must be capable of reacting selectively with the respective receptive substance. There are chemical difficulties in the path of any blunt application of this view, but what little we know of protoplasmic processes would appear to favor the idea that the essential stimulus-reaction may well be removed by several chemical steps from the very first union of taste-evoking solute with the surface of the receptor cell, and in this way the theory might still be saved. In its support may be cited such experiments as the following, due to Jefimoff in Lasareff's laboratory. Gustatory adaptation was produced by sugar solution on the one hand, and in a similar way by saccharine. Presumably both act upon the same "receptor substance" to produce a sweet taste. The sensitivity E being reduced to the same level in each case, the recovery of excitability is then followed, as a function of the elapsed time, through determinations of the concentration of sugar required to excite. We are not told of the precautions taken to avoid adaptation during the course of the tests. Results are given in Table 8.

TABLE 8

Values of gustatory sensitivity E, with sugar as test substance, are obtained during recovery of excitability following gustatory exhaustion (A) by sugar solution, (B) by saccharine solution. E is also given as calculated from

$$E = 11.5\ (1 - e^{-0.16t}),$$

where t = time for recovery adaptation (seconds).

Adapted by					
(A) Sugar			(B) Saccharine		
t secs.	E obs.	E calc.	t secs.	E obs.	E calc.
4.5	6.0	5.9	5	6.0	6.4
7.5	8.4	8.1	9	8.0	8.8
16.5	11.1	10.6	19	10.0	11.0
48.8	12.5	11.5	37	11.7	11.5
			40	12.1	11.5

While agreements between observed and computed changes in sensitivity are not impressive, and we are not given the statistical weights of the figures, there is some indication that recovery of sensitivity to a sugar solution follows the same time-course after sensory exhaustion by sugar as after similar exhaustion by saccharine. This result would be expected if both sugar and saccharine in provoking the sweet taste entered into an identical reaction-complex (again using "reaction" with chemical connotation). In this way the curious similarity of effect as due to two such molecularly different substances as sugar and saccharine can be understood.

It cannot be held, however, that there is self-evident necessity in the assumption of four separate reaction-complexes corresponding to the elementary taste qualities. The problem is like that of color perception, where it may not be required to assume three color-receptive retinal substances. But into this aspect of the problem, the poverty of relevant data does not permit us to go. It becomes even more embarrassing in the treatment of smell.

REFERENCES

1. Abbott, C. E. Experimental data on the olfactory sense of Coleoptera, with special reference to the Necrophori. *Ann. Entom. Soc. Amer.*, 1927, **20**, 207-216.
2. Adrian, E. D. The basis of sensation. New York: Norton, 1928. Pp. 122.

3. Allison, V. C., & Katz, S. H. An investigation of stenches and odors for industrial purposes. *J. Indus. Eng. Chem.*, 1919, **11**, 336-338.

4. Arey, L. B., & Crozier, W. J. The sensory reactions of *Chiton*. *J. Exper. Zoöl.*, 1919, **29**, 157-260.

5. Arnstein, C. Die Methylenblaufärbung als histologische Methode. *Anat. Anz.*, 1887, **2**, 125-135.

6. Asai, T. Untersuchungen über die Struktur der Riechorgane bei *Mustelus laevis*. *Anat. Hefte, Arb.*, 1913, **49**, 441-521.

7. Backman, E. Note sur la puissance des odeurs et leur solubilité dans l'eau et dans l'huile. *J. de physiol. et de path. gén.*, 1917, **17**, 1-4.

8. Barral, —, & Kanc, —. La chimie des édulcorants. *Rev. scient.*, 1918, **56**, 712.

9. Becker, C. T., & Herzog, R. O. Zur Kenntnis des Geschmackes. *Zsch. f. physiol. Chem.*, 1907, **52**, 496-505.

10. Beer, T., Bethe, A., & von Uexküll, J. Vorschläge zu einer objektivierenden Nomenklatur in den Physiologie des Nervensystems. *Biol. Zentbl.*, 1899, **19**, 517-521.

11. Broman, I. Om Jacobsonska Organets konstruktion om funktion. *Lunds. Univ. Arsskrift, N. F.*, Avd. 2, 1918, **14**. Pp. 40.

12. Brünn, A. Beiträge zur mikroskopischen Anatomie der menschlichen Nasenhöhle. *Arch. f. mikr. Anat.*, 1892, **39**, 632-651.

13. Cohn, G. Geschmack und Konstitution bei organischen Verbindungen. *Samml. chem. u. chem.-techn. Vorträge*, 1918, **22**, 1-100.

14. Cowdry, E. V. [Ed.] Special cytology. (2 vols.). New York, 1926. Pp. 1376.

15. Crozier, W. J. The sensory reactions of *Holothuria surinamensis* Ludwig. *Zoöl. Jahrb., Abt. f. Physiol.*, 1915, **35**, 233-297.

16. ————. Tonic antagonism in sensory stimulation. *Amer. J. Physiol.*, 1915, **39**, 297-302.

17. ————. Regarding the existence of the "common chemical sense" in vertebrates. *J. Comp. Neur.*, 1916, **26**, 1-8.

18. ————. The taste of acids. *J. Comp. Neur.*, 1916, **26**, 453-462.

19. ————. Cell penetration by acids. I. *J. Biol. Chem.*, 1916, **24**, 255-279.

20. ————. Cell penetration by acids. II. Further observations on the blue pigment of *Chromodoris zebra*. *J. Biol. Chem.*, 1916, **26**, 217-223.

21. ————. Cell penetration by acids: III. Data on some additional acids. *J. Biol. Chem.*, 1916, **26**, 225-230.

22. ————. Cell penetration by acids. IV. Note on penetration of phosphoric acid. *J. Biol. Chem.*, 1918, **33**, 463-470.

23. ————. Sensory activation by alkalies. *Amer. J. Physiol.*, 1918, **45**, 315-322.

24. ————. Sensory activation by acids. *Amer. J. Physiol.*, 1918, **45**, 323-341.

25. ————. Cell penetration by acids. V. Note on the estimation of permeability changes. *J. Gen. Physiol.*, 1922, **4**, 723-731.

26. ————. Cell penetration by acids. VI. The chloroacetic acids. *J. Gen. Physiol.*, 1922, **5**, 65-79.

27. Cushing, H. The taste fibers and their independence of the N. trigeminus. *Johns Hopkins Hosp. Bull.*, 1903, **14**, 71-78.

28. Ehrlich, P. Ueber die Methylenblaureaction der lebenden Nervensubstanz. *Dtsch. med. Woch.*, 1886, **12**, 49-52.

29. Fischer, E., & Penzoldt, F. Ueber die Empfindlichkeit des Geruchssinnes. *Sitzber. phys.-med. Soc. Erlangen*, 1886, **18**, 7-10.

30. Frey, M. von. Der Laugige Geruck. *Pflüg. Arch f. d. ges. Physiol.*, 1910, **136**, 275-281.

31. Frisch, K. von. Ueber den Geruchsinn den Biene und seine blütenbiologische Bedeutung. *Zoöl. Jahrb., Abt. f. Physiol.*, 1919, **37**, 1-238.

32. ————. Ueber den Sitz des Geruchsinnes bei Insecten. *Zoöl. Jahrb., Abt. f. Physiol.*, 1921, **38**, 449-516.

33. ————. Versuche über den Geschmackssinn der Bienen. *Naturwiss.*, 1927, **15**, 1-20.

34. ————. Vergleichende Physiologie des Geruchs- und Geschmackssinnes. 1928.

35. Fröhlich, R. Ueber einige Modificationen des Geruchssinnes. *Sitzber. Akad. wiss. Wien, math.-nat. Cl.*, 1851, **6**, 322-338.

36. Gellhorn, E. Das Permeabilitätsproblem. Berlin, 1929. Pp. x+441.

37. Giersberg, H. Eck. Untersuchungen über den Geschmackssinn der Molche. *Zsch. f. vergl. Physiol.*, 1926, **3**, 377-388.

38. Hänig, D. P. Zur Psychophysik des Geschmackssinnes. *Phil. Stud.*, 1901, **17**, 576-623.

39. Hanström, B. Einige Experimente und Reflexionen über Geruch, Geschmack und der allgemeinen chemischen Sinn. *Zsch. f. vergl. Physiol.*, 1926, **4**, 528-544.

40. Hartwell, R. A. A study of the olfactory sense of termites. *Ann. Entom. Soc. Amer.*, 1924, **17**, 131-160.

41. Heidenhain, M. Ueber die Sinnesfelder und die Geschmacksknospen der Papilla foliata des Kaninchens. *Arch. f. mikr. Anat.*, 1914, **85**, 365-479.

42. Heidrich, F. Die Zahl und die Dimension der Geschmacksknospen der Papilla vallata des Menschen in den verschiedenen Lebensaltern. *Nachr. Gesell. Wiss. Göttingen, math. phys. Cl.*, 1906, 54-64.

43. Henning, H. Physiologie und Psychologie des Geschmacks. *Erg. d. Physiol.*, 1921, **19**, 1-78.

44. ————. Der Geruch. (2nd ed.) Leipzig, 1924.

45. Herlitzka, A. Sul "sapore metollico," sulla sensazione astringente e sul sapore dei sali. *Arch. di fisiol.*, 1908, **5**, 217-242.

46. Herrick, C. J. The organ and sense of taste in fishes. *Bull. U. S. Fish. Comm.*, 1903, **22**, 237-272.

47. Hoeber, R. Physikalische Chemie der Zelle und der Gewebe. (6th ed.) Leipzig: Engelmann, 1926.

48. Hollingworth, H. L., & Poffenberger, A. T. The sense of taste. New York: Moffat, Yard, 1917.

49. Hopkins, A. E. The olfactory receptors in vertebrates. *J. Comp. Neur.*, 1926, **41**, 253-290.

50. Jagodowski, K. P. Zur Frage nach der Endigung der Geruchsnerven bei den Knochenfischen. *Anat. Anz.*, 1901, **19**, 257-267.

51. Kahlenberg, L. The action of solutions on the sense of taste. *Bull. Univ. Wis. Sci. Ser.*, 1898, **2**, 1-31.

52. Kiesow, F. Contribution à la psycho-physiologie de la cavité buccale. *Arch. ital. de biol.*, 1898, **30**, 377-398.

53. ————. Ein Beitrag zur Frage nach den Reaktionszeiten der Geschmacksempfindungen. *Zsch. f. Psychol. u. Physiol. d. Sinnesorg.*, 1903, **33**, 453-461.

54. Kremer, J. H. Adsorption de matières odorantes et de narcotique odorants par les lipoïdes. *Arch. néerl. de physiol.*, 1917, **1**, 715-725.

55. Larguier des Bancels, L. Le goût et l'odorat. Paris, 1912. Pp. vii+94.

56. Lasareff, P. Théorie ionique de l'excitation des tissus vivants. *Coll. monog. scient. étrang.*, 1928. Pp. iii+240.

57. Liebermann, A. Correlation zwischen den antennalen Geruchsorganen und der Biologie der Musciden. *Zsch. f. Morphol. u. Oekol. d. Tiere,* 1925, **5**, 1-97.

58. Lillie, R. S. Protoplasmic action and nervous action. Chicago, 1923.

59. Lovén, C. Bidrag till Kännedomen om tungans smakpapiller. 1867.

60. Lucas, K. The evolution of animal function. II. *Sci. Prog.,* 1909, **4**, 321-331.

61. Matthes, E. Das Geruchsvermögen von Triton, beim Aufenthalt unter Wasser. *Zsch. f. vergl. Psychol.,* 1924, **1**, 57-83.

62. ————. Das Geruchsvermögen von Triton beim Aufenthalt an Land. *Zsch. f. vergl. Physiol.,* 1924, **1**, 590-606.

63. May, R. M. The relation of nerves to degenerating and regenerating taste buds. *J. Exper. Zoöl.,* 1925, **42**, 371—.

64. McIndoo, N. E. The olfactory sense of the honey-bee. *J. Exper. Zoöl.,* 1914, **16**, 265-346.

65. ————. Senses of the cotton boll weevil: an attempt to explain how plants attract insects by smell. *J. Agric. Res.,* 1926, **33**, 1095-1141.

66. ————. Tropisms and sense organs of Lepidoptera. *Smithson. Miscel. Collection,* 1929, **81**. Pp. 59.

67. Minnich, D. E. The olfactory sense of the cabbage butterfly, *Pieris rapae* Linn. *J. Exper. Zoöl.,* 1924, **39**, 339-356.

68. ————. The chemical sensitivity of the tarsi of certain muscid flies. *Biol. Bull.,* 1926, **51**, 166-178.

69. ————. The organs of taste on the proboscis of the blowfly, *Phormia regina* Meigen. *Anat. Rec.,* 1926, **34**, 126.

70. Nagel, W. Vergleichende physiologische und anatomische Untersuchungen über den Geruchs- und Geschmackssinn und ihre Organe. *Bibl. Zoöl.,* 1894. No. 18.

71. Oehrwall, H. Untersuchungen über den Geschmackssinn. *Skand. Arch. f. Physiol.,* 1891, **2**, 1-69.

72. ————. Die Modalitäts- und Qualitätsbegriffe in der Sinnesphysiologie und deren Bedeutung. *Skand. Arch. f. Physiol.,* 1901, **11**, 245-277.

73. Oertly, —, & Myers, —. *J. Amer. Chem. Soc.,* 1919, **41**, 855.

74. Olmsted, J. M. D. The nerve as a formative influence in the development of taste-buds. *J. Comp. Neur.,* 1920, **31**, 465.

75. ————. The results of cutting the seventh cranial nerve in *Amiurus nebulosus. J. Exper. Zoöl.,* 1920, **31**, 369.

76. ————. Effects of cutting the lingual nerve of the dog. *J. Comp. Neur.,* 1921, **33**, 149.

77. ————. Taste fibers and the chords tympani nerve. *J. Comp. Neur.,* 1922, **34**, 337.

78. Osterhout, W. J. V. The nature of mechanical stimulation. *Proc. Nat. Acad. Sci.,* 1916, **2**, 237-239.

79. ————. Injury, recovery, and death in relation to conductivity and permeability. Philadelphia, London: Lippincott, 1923.

80. Osterhout, W. J. V., & Dorcas, M. J. The penetration of CO_2 into living protoplasm. *J. Gen. Physiol.,* 1925-1926, **9**, 225.

81. Parker, G. H. The sense of taste in fishes. *Science,* 1908, **27**, 453—.

82. ————. The sensory reactions of *Amphioxus. Proc. Amer. Acad. Arts & Sci.,* 1908, **53**, 415-455.

83. ————. The reactions of sponges, with a consideration of the origin of the nervous system. *J. Exper. Zoöl.,* 1910, **8**, 1-41.

84. ————. The olfactory reactions in fishes. *J. Exper. Zoöl.,* 1910, **8**, 535-542.

85. ————. The relation of smell, taste, and the common chemical sense in vertebrates. *J. Acad. Nat. Sci. Philadelphia,* 1912, **15**, 221-234.

86. ————. The elmentary nervous system. Philadelphia, London, 1919.

87. ————. Smell, taste, and allied senses in the vertebrates. Philadelphia, London, 1922. Pp. 192.

88. ————. The neurofibril hypothesis. *Quar. Rev. Biol.,* 1929, 155-178.

89. Parker, G. H., & van Heusen, A. P. The responses of the catfish, *Amiurus nebulosus,* to metallic and non-metallic rods. *Amer. J. Physiol.,* 1917, **44**, 405-420.

90. Parker, G. H., & Stabler, E. M. On certain distinctions between taste and smell. *Amer. J. Physiol.,* 1913, **32**, 230-240.

91. Paul, T. Investigation of acid taste. *Zsch. Electrochem.,* 1922, **28**, 435-446.

92. Read, E. A. A contribution to the knowledge of the olfactory apparatus in dog, cat, and man. *Amer. J. Anat.,* 1908, **8**, 17-47.

93. Renqvist, —. *Skand. Arch. f. Physiol.,* 1918, **40**, 124.

94. Retzius, G. Die Nervenendigungen in dem Geschmacksorgan der Säugetiere und Amphibien. *Biol. Untersuch., N. F.,* 1892, **4**, 19-32.

95. ————. Die Riechzellen der Ophidier. *Biol. Untersuch., N. F.,* 1894, **6**, 48-51.

96. Reusch, B., & Eisentraut, M. Experimentelle Untersuchungen über den Geschmackssinn der Reptilien. *Zsch. f. vergl. Physiol.,* 1927, **5**, 607-612.

97. Richards, T. W. The relation of the taste of acids to their degree of dissociation. *Amer. Chem. J.,* 1898, **20**, 121-126.

98. Rubaschkin, W. Ueber die Beziehungen des Nervus trigeminus zur Riechschleimhaut. *Anat. Anz.,* 1903, **22**, 407-415.

99. Schaeffer, J. P. The nose, paranasal sinuses, nasolacrimal passage ways, and olfactory in man. Philadelphia, 1920. Pp. 370.

100. Schwalbe, G. Das epithel der Papillae vallatae. *Arch. f. mikr. Anat.,* 1867, **3**, 504-508.

101. Sheldon, R. E. The reactions of the dogfish to chemical stimuli. *J. Comp. Neur. & Psychol.,* 1909, **19**, 273-311.

102. Strieck, R. Untersuchungen über den Geruchs- und Geschmackssinn der Ellritze. *Zsch. f. vergl. Physiol.,* 1924, **2**, 122-153.

103. Taylor, N. W. Acid penetration into living tissues. *J. Gen. Physiol.,* 1928, **11**, 207-221.

104. ————. A physico-chemical theory of sweet and bitter taste excitation based on the properties of the plasma membrane. *Protoplasma,* 1928, **4**, 1-17.

105. Van der Stricht, O. Le neuro-épithélium olfactif et sa membrane limitante interne. *Mém. Acad. roy. méd. Belgique,* 1909, **20**. Pp. 45.

106. Veress, E. Ueber die Reizung des Riechorgans durch directe Einwirkung riechende Flussigkeiten. *Arch. f. d. ges. Physiol.,* 1903, **95**, 368-408.

107. Verlaine, L. Le déterminisme du déroulement de la trompe et la physiologie du goût chez les Lépidotères (*Pieris Lapae* Linn.) *Ann. et bull. Soc. entom. Belgique,* 1927, **67**, 149-182.

108. Wacker, F. Beiträge zur Kenntniss der antennalan Sinnesorgane der Hymenopteren. *Zsch. f. Morphol. u. Oekol. d. Tiere.,* 1925, **4**, 739-812.

109. Weber, E. H. Ueber den Einfluss der Erwärmung und Erkältung der Nerven auf ihr Leitungsvermögen. *Arch. Anat. Physiol.,* 1847, 342-356.

110. Wheeler, W. M. Social life among the insects. New York, 1928.

111. Woodrow, H., & Karpman, B. A new olfactometric technique and some results. *J. Exper. Psychol.,* 1917, **2**, 431-447.

112. Wundt, W. Grundzüge der physiologischen Psychologie. (3rd ed., vol. 1.) 1887.

113. Zwaardemaker, H. L'odorat. (Encycl. scient.) Paris, 1925. Pp. 305.

CHAPTER 9

THE SENSE OF FEELING

JOHN PAUL NAFE
Clark University

The sense of feeling includes all those psychological experiences that are "felt" in the common meaning of that word and its study must include: (1) a description of the observable psychological phenomena, (2) a description of the neurological conditions under which such experiences occur, and (3) a description of the environmental (stimulus) conditions involved. We refer to the first of the above divisions of the subject as the "psychology" of feeling and to the second and third as the "theory" of feeling because psychologists "explain" by relating psychological experience with its neurological and environmental conditions.

The organ of this sense comprises practically the whole body, the afferent fibers being so distributed that the periphery of the body contains a mesh of fibers finely woven and many strands thick. The interior of the body also, in all its parts, is richly supplied. The receptors and nerve-endings are scattered unevenly but in great numbers throughout the body. They include all the receptors in the body except those of the special senses of vision, audition, olfaction, and gustation, and, perhaps, also those of the non-acoustic labyrinths concerned with equilibrium, etc.

The stimulating objects or conditions for feeling may be, as is often stated, mechanical, chemical, thermal, or electrical, but the action of such stimuli is not to be considered as immediate. Their occurrences are accompanied by corresponding changes in the conditions within the tissues and, as a result of such changes within the receptors, nervous impulses are initiated. We speak of such changes as "discharging" the receptors but, since we know so little about the action of receptors, the term "discharge" has little real significance. All such changes within the tissues may be thought of as tending toward an establishment of equilibrium conditions, equilibrium being broken down by the introduction of each new set of forces in the environment. Within such a system, unless the change is great enough, no impulse is initiated and only adaptation occurs.[1]

Stimulation requires time, and it makes no difference in principle whether the time be relatively long or short. The processes involved in the establishment of equilibrium conditions begin very soon after the change in environmental conditions has begun and, within possible limits

---

[1] This is not a proper place for a full discussion of theories of stimulation. Cf. Chapter 3, this book. The current theories of mechanical and thermal stimulation have little to support them, and it is becoming apparent that the immediate stimulus for "felt experience" is of the nature described above. It probably consists of shifts in the ionic concentrations within or on the surface of cells.

(chronaxie, etc.), the greater the amount of change within a unit of time, the greater the frequency of impulses initiated in the nerve-fiber.

If we consider "adaptation" to be the establishment of equilibrium conditions, then the "adaptation-time" is the time required to establish such conditions. It varies with the intensity and duration of stimulation. Adaptation-time may be considered to include some time after the removal of the stimulus, or such a change may be regarded as a new set of conditions and consequently another occurrence. It makes little difference which of these views we adopt for, with such a change, one of two things may occur: (1) There may be an accumulated amount of change which continues the stimulation in the same direction for a period of time after the removal of the stimulus, in which case we should experience a positive after-image or after-sensation. (2) A change in the opposite direction may take place within the receptor, in which case we should experience a positive or negative after-image, depending upon other conditions of the discharging receptors. It is doubtful whether, in feeling, we ever experience a negative after-image.

Adaptation-time, in this sense, is synonymous with "duration." If the stimulus is applied continuously, the duration of the experience and adaptation time will be equal. If the stimulus is removed before complete adaptation, the duration of the experience plus the duration of the after-image will equal the adaptation-time. If any other change of the stimulus is effected, a new set of conditions is set up which then proceeds to equilibrium under the same principle. Ordinarily "adaptation" refers to the duration of experience when the application of the stimulus is continuous, and "duration" refers to the continuance of experience either as limited by the presence of the stimulus or as limited by attention.

"Subliminal stimulation" refers to that change which is not sufficiently intense to be accompanied by the initiation of impulses. In such a case, adaptation occurs and, if the amount of stimulation is sufficiently low and the time long, changes of considerable magnitude may occur within the receptor without being accompanied by the initiation of impulses. A succession of subliminal stimulations may finally result in a series of impulses, a phenomenon known as "summation."

Under usual conditions of stimulation a series of impulses is aroused in a number of fibers, and little is known about single nervous impulses and little about impulses aroused in a single nerve-fiber. Stimulation, even of comparatively small areas, is usually accompanied by the initiation of impulses over a large number of fibers.

The specific fibers activated, under different conditions of stimulation, vary and the "discharges of nervous impulses" may vary from each other from time to time in the following respects:

1) In the frequency of the impulses aroused in the individual fibers.

2) In duration or the length of time during which the impulses continue in each fiber.

3) In the size of the area or extent over which fibers are activated.

4) In the relative number of fibers activated within such an area (density).[2]

The first two variable aspects mentioned, "frequency" and "duration," determine the *temporal* relations of impulses set up over the individual fibers: rate and total time, and the two together determine the total number of impulses aroused for each fiber. The second two variable aspects determine the *spatial* relations of impulses set up over a number of fibers: relative density and total extent, and the two together determine the total number of fibers involved in each occurrence. All four together determine the whole number of impulses and the "pattern" or "arrangement" of any given "discharge" as a whole, and great variation is possible between separate occurrences of such groups of impulses or "patterns of discharges."

Variation in the separable aspects of neural discharges described above have very definite relations both with the conditions of stimulation and with the type of conscious experience aroused.

1) Frequency of impulses, as Adrian and Zotterman (4) have shown, is directly proportional to the amount of stimulation. Psychologically, variation in frequency of impulses is directly proportional to the intensity of experience as felt.

2) The duration of series of impulses is directly proportional to the amount of change set up within the receptors and the length of time such change continues. Psychologically, it is correlated with the duration of experience, which may be subdivided to distinguish after-image or after-sensation from the rest of the experience, or the latter phase may be regarded as a separate experience. This view of "duration" is probably common to all theories although it is variously expressed.

3) The size of the area over which the discharge occurs varies with the size of the stimulus, but it also varies with the intensity of stimulation because of the pliable nature of living tissues. This is correlated on the psychological side with the felt extensity.

4) Stimulation with moderately intense stimuli does not set up impulses in all the fibers originating within an area so stimulated and, of those fibers which are activated, all do not begin to function at the same time. With increased intensity of stimulation more fibers respond and those which responded to the former conditions now respond with a greater frequency of impulses. We refer to the relative number of fibers activated within such an area as the "density" of discharge. As the intensity of stimulation continues to increase, the less irritable fibers increase in irritability more rapidly than the more irritable so that, at higher intensities of stimulation, there is a tendency for the impulses to begin more

---

[2] The list is oversimplified. It has been shown that in some nerves there is a difference between individual fibers in the rate of conduction of impulses. What such variations may mean functionally we do not know (22). Records of series of discharges may vary in the amplitude of the waves and some records show rhythms of discharge. The functional importance of these variations is not known and at present they play no part in the theory of feeling.

nearly together (22, p. 564). Variations in density are thus ordinarily correlated with variation in intensity of stimulation, but we may set up conditions to increase density alone. If we stimulate with objects such as grills, a finer grill will be felt as "denser" than a coarser one without necessarily any increase in other variable aspects of experience.

In any normal occurrence, variation in the *intensity* of stimulation is accompanied by directly proportional variations of the frequency, duration, extensity, and density of impulses; variation in the *form* of such stimulation is accompanied by variations in some of the four aspects independently of each other, e.g., a change in the *size* of the stimulated area, other variable conditions remaining constant, will affect only felt *extensity*.

The Quantitative Theory of Feeling

Variations in the frequency of impulses are correlated on the psychological side with changes in felt intensity, which vary between the relatively weak or "dull" experiences of the pressure type to the relatively strong, or "bright," or "sharp" experiences of contact, pain, tickle, etc.[3] This means that the difference between "pressure experiences" and "pain experiences" is primarily a matter of intensity, and reference to the section on Sensitivity to Temperature Differences will show that experiences of warmth and cold are not exceptions to this general principle. The concept of "quality," either as an independently variable aspect of conscious experience or as an explanatory neurological principle, is eliminated in this theory. There is only one "quality" of felt experience, hence the term loses its earlier significance as regards qualitative series and qualitative differences and in theory requires no specialized receptors or specialized cortical areas for identification.

Variations in duration, extensity, and density of discharge are, for this theory as for others, correlated with felt duration, extensity, and density.

The cortical locus of the central connection of particular fibers serves as the criterion for localization, and localization must be learned. Such learning consists of conditioning the particular responses involved to occurrences within the central nervous system which, themselves, do not rise to consciousness.

The particular experiences, e.g., wet, ache, pressure, etc., vary from each other in the "pattern of discharges," i.e., in the number of impulses involved and their temporal and spatial relations. Similar stimulus objects will, from time to time, set up similar conditions within the tissues, and these will be accompanied by similar psychological experiences. The fact that some such similar experiences recur with sufficient frequency to be named indicates nothing as to their ultimate simplicity, nor does the ability or inability of an observer to analyze such an experience constitute any evidence as to the neurological mechanism by which it is aroused.

Further complications of experience are occasioned by reflex activities arising either from felt experiences themselves or from other sensory experiences. Such activities are of two very different sorts:

---

[3] Intensity or brightness is not equivalent to "painfulness" or the amount an experience "hurts" although they vary together. Cf. section on Pain.

1) Activities of muscular and glandular systems which are followed by the arousal of afferent impulses and consequent felt experience. When an object comes into contact with the skin, one may see, under a binocular microscope, changes occurring within the tissues. These visible changes may consist of movement of skin and dilation of superficial capillaries, the red blood showing near the surface, and, if the contacts are repeated to the extent of chafing, the changes become apparent without magnification as inflammation and change in texture of the skin. These activities, however, are not limited to the small changes near the point of stimulation but may include shudders, chills, thrills, laughter, twitching, weeping, quivers, nausea, perspiring, temporary blindness, goose-flesh, changes in tonus, etc., or even complicated, coordinated activity of muscle systems. Such reflexly aroused activities as these in turn arouse afferent impulses and consequent feelings which in less intensive experiences tend to modify the existing pattern, but in more intensive experiences tend to be singled out as an affective or emotional or other accompaniment of the more directly aroused part of the experience. If "affective tone" ultimately refers to nothing more than the feelings aroused by such reflexes, it is probably true that every experience is accompanied by an "affective tone" although with experiences of relatively low intensity the induced activities are very slight in their conscious consequences.

2) Consciousness itself is reflexly aroused. The more intensive a stimulation the more likely is it to be accompanied by a conscious experience. Series of impulses of slow frequency are not likely to be accompanied by consciousness at all. In these conditions of consciousness lies the answer to the question whether all receptors are also sensory end-organs. Potentially they are, but actually they may never become so, depending upon the general conditions correlated with consciousness, which conditions, so far as they are known, are expressed in the "laws of attention."[4]

Each of these reflexly aroused experiences modifies the existing experience, and the number and vigor of such reflexes, aroused with felt experiences, vary with the intensity and form of stimulation. Regardless of the number and amount of its complications, the neural excitation remains a temporal and spatial "arrangement" or "pattern" of a number of impulses and as such may be accompanied by a related conscious experience. In all these complications of "patterns of discharges" there are no new elements. Familiar patterns are named, and if an experience arises for which we haven't names enough, we supplement our description with qualifying adjectives. After some unusual experience, such as a surgical operation, we may even launch into a full description, but all the terms used, descriptive of felt experience, imply variations of patterns and are described in terms of intensity, density, extensity, duration, and localization or collocations of such aspects within the experience as a whole.

---

[4] Some authors posit "attensity" as an independent variable of experience comparable with "intensity," etc. Such a view has received scant support (54, 55).

The aspects of this theory which apparently need explication or proof are: (1) those having to do with the variation of the pressure-pain series of experiences with frequency of impulses and amount of stimulation rather than with the specialized receptors and the cortical locus of the central connection of the particular fiber stimulated (the correlate for "quality" under the older theory); (2) the reconciliation of experiences accompanying temperature differences with the theory. This second point will be clarified in the section entitled "Sensitivity to Temperature Differences."

The necessity for evidence upon the first point arises more from the tradition of "qualities" and Müller's so-called "law of specific nerve energies" than from any fact known to controvert it (38). The enunciation of Müller's law was accompanied by no factual demonstration nor have the intervening years furnished any convincing proofs. As a theory it has resulted in a confusion of irreconcilable facts and dogmas, and its present importance is due only to its wide acceptance, not to its factual basis nor to its systematic applicability. The question here at issue is whether pressure and pain experiences depend upon the variations of impulses over more than one system of nerves or over a single system of nerves. The evidence for the former possibility is presented in the section upon The Qualitative Theories, and that for the latter possibility is presented below in the form of evidence tending to show a correlation between (*a*) the amount of stimulation, (*b*) the frequency of impulses, and (*c*) the accompanying feelings of pressure and pain types.[5]

The correlation between amount of stimulation and frequency of impulse is well established if we put the proof in the language of the qualitative theory. For pressure it is true, for pain it is true, but what has always been accepted, without supporting facts, is the assumption that there are "pressure nerves" and "pain nerves." Upon the ground of economy, those who adopt this position must carry the burden of proof.

Adrian and Zotterman have reported a series of experiments in which records of afferent impulses were made.[6] Some of these records are of impulses from receptors in muscle tissue and some from the cutaneous receptors. Many of them do not show great differences in frequencies of impulses resulting from different types of stimulation, but in working with impulses aroused by a *single receptor* within a muscle it was found that an increase in the amount of stimulation is followed by an increase in the frequency of the impulses (4). This is in accordance with all other pertinent findings, and we must generalize from this one *known* relation. The generalization is subject to any evidence produced tending to show exceptions but it is a presumption which we cannot dismiss in any instance without direct proof to the contrary.

---

[5] The evidence from psychology is equivocal, apparently fitting either hypothesis. Series of experiences ranging from pressure to pain may be set up by increasing only the intensity of stimulation, but such facts may be accounted for under either theory (42). Insensitivity of certain areas to experiences of one kind or another may also be accounted for under either theory because we have no adequate theory of immediate stimulation, i.e., there may lack a specific type of receptor or there may lack the proper conditions within the tissues.

[6] The neurological evidence in this section is taken largely from the reports of these investigators (1, 2, 4, 5). Some of the work is re-presented in *The Basis of Sensation* (3). Familiarity with the work of Adrian and Zotterman is necessary to an understanding of the present status of the problem.

In the interpretation of the records of many experiments in this series, a difficulty arises in that the number of fibers is not known, and an increase or decrease in the record of frequency of impulses may be due to variation in the number of fibers activated or may be due to variation in the frequency of impulses in the individual fibers, or both. Interpretation of results in regard to frequency is further hindered by the similarity of the records of the first moment of stimulation, a confusion due to the lack of an adequate theory of stimulation. Small currents of electricity aroused by the adjustment of differences in potential between living tissue and foreign objects may account for some of the similarities in the records in this respect, especially when the nature of the stimulus is such as to permit the establishment of a focal point. Experiences aroused with brief or intermittent punctiform stimulation may be very intensive, although fleeting and not necessarily painful, e.g., tickle, itch, and experiences of like character which are attention-demanding and reaction-provoking to a degree comparable with pain itself. There is also a possible difficulty due to the fact that in such measurements the actual amount of stimulation *per unit area* is not always calculated. The point of a needle resting upon the skin with the weight of a few grams is equal, in pressure per unit area, to several hundred grams over an area of 1 sq. cm. After making allowance for these factors, the records of this series of experiments become more significant.

The stimulation of the tissues of a frog with a needle weighted in one series with 43 and in another series with 4 gms., except for the first period (.50 sec.), shows that the less intense stimulus consistently produces a lower frequency of impulses, and the conditions of stimulation seem to be comparable (2, esp. p. 47)

In another experiment the limb of the preparation (cat) containing the part of the skin under observation rested upon a table (1, p. 65 ff.). The record showed a rapid series of oscillations due, presumably, according to Adrian, to the pressure of the leg on the table and the movement of hairs by air currents. These oscillations "increase in frequency and amplitude if the skin of the foot is nipped by a pair of artery forceps or is pricked by a pin." "The increase in frequency may be relatively small if the frequency is already high in the resting limb. . . .[but]. . . . there is no indication that the additional impulses set up by a painful stimulus differ much from those with the limb undisturbed, which are presumably due to non-painful stimuli."[7]

---

[7] An objection may be raised to the direct comparison between the sensitivity of a decerebrate preparation and that of a normal human being, but we offer such a comparison as the best available evidence.

Adrian reports that the needle used in certain of these experiments, when weighted with 3 gms. and applied to a finger, produced a mild experience of contact, but when weighted with 43 gms. produced a prick. When applied to a preparation capable of reacting, the needle weighted with 43 gms. produced decided flexion. It is significant that stimulation with "painful" stimuli, where the record is clear, results in a higher frequency than with "non-painful" stimuli while no such record shows a lower frequency.

There is no indication of overlapping in the frequencies characteristic of pressure and pain experiences where the records are clear nor does Adrian find other differences characteristic of the records for the two types of experience. There is no difference shown in the duration of single impulses or in the grouping of impulses. Bearing upon this point, in the discussion of his results, Adrian says in part: "The close likeness from the records of the vagus, internal saphenous and the frog's sciatic suggests either that all the sensory fibers concerned are very much alike, or that the characteristic record is an artifact with oscillations determined by the recording instrument. This possibility has been discussed already and rejected. . ." The most noticeable difference between the records produced with "painful" stimuli and "non-painful" stimuli in this series of experiments, and the only difference accepted by Adrian as significant, is that between adaptation-times, which is much greater for "painful" than for "non-painful" stimuli.[8]

The records, in regard to frequency of impulse, are often equivocal because of the overlapping of recorded impulses. The differences apparently being not great, Adrian decides against the probability of any *significant* differences in frequency. This is placing Müller's doctrine ahead of his own experimental findings cited above, in which he found positive evidence of a correlation between the amount of stimulation and frequency of impulse. The presumption in all such records must be that there *is* a significant difference in frequency unless there is positive evidence to the contrary.

It becomes evident that the simplest explanation of these facts must posit a series of experiences ranging from the less intensive or dull of pressure to the more intensive or bright of pain and contact, which vary directly with the frequency of impulses and the amount of stimulation.

THE QUALITATIVE THEORIES

The older theories of feeling, to which we shall refer as the qualitative theories, in distinction to the quantitative theory treated in the last section, appear in a variety of forms. The principal differences between the two theories lie in the acceptance and rejection of the concept of "qualities" as an explanatory principle in feeling; and in the failure on the part of the older theories to account satisfactorily for localization. There are many other differences between these two types of theory and there are differences among the qualitative theories but they are not essential to either theory.

---

[8] Cf. section on Pain.

Localization. The doctrine of "local sign" originated with Lotze (35). It hypothecates, within each experience, a *qualitative* difference of which one is not conscious but which designates the position from which the impulse arises. The objections to this doctrine are as follows:

1) "Qualitative differences of which one is not conscious" is a confusion of terms. Under the older theory, qualities are aspects of conscious experience and the position is essential to their theory. They may not, then, posit conscious aspects of which one is not conscious.

2) In the qualitative theories every neurological principle of variation within the limits of our present knowledge is used in accounting for other known differences so that the theory not only hypothecates a quality of which we know nothing but correlates this quality with some principle of neurology yet to be discovered.

3) Localization has been shown to be a learned process (23, 25, 37, 46, 48, 49). If based upon a qualitative variation it must be directly experienced.

Qualities. Qualitative variation is based upon the following considerations:

1) A common-sense knowledge of different experiences and a general acceptance of Müller's law of specific nerve energies.

2) Evidence of a relation between the sensitive areas of the skin and the histologically differentiated receptors of the skin.

3) The fact that some of these qualities may be eliminated, the rest persisting, as shown by:

a) the application or injection of drugs,
b) nerve section,
c) lesions in the spinal cord.

1) Müller's law, as commonly interpreted, may be expressed in the language of today as follows: The specific [conscious] accompaniment of sensory excitation is correlated with the specific response in the particular [cortical] structure in which the nerve-fiber ends. Practically it also involves the specificity of the receptors.

This means that "qualitative" differences are correlated with different fibers, with different cortical areas, and with functionally specialized receptors excited by a specific (adequate) stimulus. Since there are not enough different kinds of receptors to posit a separate nerve system for every kind of felt experience, some of the qualities are called "simple" and others "patterns" or "complexes" of these more simple ones. There is no known criterion for simple qualities except the impossibility of psychological analysis, and there is no agreement as to which experiences are unanalyzable. Some authors posit three simple qualities, others more than thirty, although convincing reasons for such categories are lacking. Titchener, attempting to determine qualitative series, produced the most understandable list (56, 42). Physiologists and neurologists, apparently without any criterion whatever, usually select pressure, pain, warm, and cold, although some treat warm and cold as one temperature sense. Few suggestions are offered as to how these "qualities" are united to form such experiences as tickle, contact, etc.

It is apparent that there are different experiences: warmth, pain, ache, wet, rough, etc. Some of these experiences obviously can be analyzed into "units" of more commonly experienced patterns and others have been shown to be so analyzable, but the ability of a psychologist or a neurologist to find, within one of these patterns, resemblances to other patterns is not proof of "complexity," and failure to find such units is no proof of "simplicity" in the sense that such experiences are mediated by separate systems of nerves.

2) The work of Blix (9) and Donaldson (21) and the supplementary work done by Goldscheider (28), von Frey (26), and others on "spots" of sensitivity upon the skin is well known. The method by which the facts have been determined will be briefly described here although many precautions and refinements of technique must be omitted.

The apparatus used consists of instruments designed for contact with the skin. For the investigation of warmth, a metal cylinder heated to a known temperature is used or, better still, a hollow cylinder through which a current of water of known temperature passes. For cold, similar instruments with lower temperatures are used. For pressure, von Frey introduced the use of hairs, which have peculiar advantages in constancy of pressure and measurability. For pain, a needle or other pointed instrument is used.

An area of the skin having been marked out, an exploration is made point by point. The results show that, with a given intensity of stimulation, experience is aroused at some points but not at others. The conclusion is drawn that where the observer reports an experience, a particular kind of organ, corresponding to the particular experience aroused, has been excited, and where the observer does not report, no such organ exists. The general conclusion is that, as an organ of touch, the skin is a mosaic corresponding to the distribution of receptors—a receptor for pressure here and one for pain there and an area between insensitive to either, etc.

The correspondence between differences in the structure of the receptors and differences in function is assumed without sufficient factual support. The structural distinctions in receptors are not accepted by all histologists, many claiming that some of these types have variants which show graded series of changes invalidating the assumption of other histologists as to essential structural differences. The evidence, as it exists, indicates that there are at least two structurally different types of endings: (1) the so-called free nerve-endings, and (2) the capsulated cell-receptors (14).

The correlation between these specialized receptors and the types of sensitivity found in the "spots" of the skin was first suggested by von Frey but without factual basis other than the existence of the "spots" of sensitivity upon the skin and the existence of the differentiated organs

(26,33). No adequate proof of the relation has since been presented. Such a postulation demands no positive disproof, and the improbability of the theory, established by a long line of extirpation experiments, practically removes it from the possibilities. (17, 18, 29, 30, 50). For every specific correlation proposed, it has been shown that the "spots" of sensitivity do not correspond in kind or number to the hypothecated specialized end-organs and in many instances do not correspond to the existence of such organs.

3) The studies upon the effect of drugs have been made for practical purposes almost exclusively. An explanation of the effects recorded has been given in terms of the current theory, but such explanations have added nothing to the probable validity of the theory itself.

Studies of the results of nerve section have been made by Henry Head (31), Trotter and Davies (57), E. G. Boring (11), and K. M. Dallenbach (19). The results of these studies are not in agreement as to the facts, and the theory of epicritic, protopathic, and deep sensitivity advanced by Head as a result of his study is not confirmed by his own facts and has not been confirmed by later work (15).

The study of spinal lesions and their effects upon sensitivity has extended over a number of years. The confused state of our knowledge of the afferent tracts in the cord has proved to be a stumbling block in this direction, and in this field it is highly probable that a reinterpretation of the facts in terms of the quantitative rather than the qualitative theory may be highly illuminating (47). The disease known as "syringomyelia," as reported, presents a case of abolition of sensitivity to warm, cold, and pain without abolishing sensitivity to pressure. This representation of the fact has led to a belief that the fibers carrying "pressure impulses" are segregated in the cord, those for heat, cold, and pain being probably closely associated but actually separate. At least some of the tests given for pressure have involved the sensitivity of receptors below the level of the skin, and it is quite possible, even probable, that the evidence for the separation of lines of conduction for the three (or more) types of sensitivity are based entirely upon a misinterpretation of the facts.

The facts, as they are now known, fail to substantiate the qualitative theory involving, as it does, specificity of receptors and cortical areas, and the attempt to understand the facts in the light of that theory has signally failed. Some of the material presented in the section upon Sensitivity to Temperature Differences constitutes further evidence, but it is apparent that the theory is not adequate to the known facts.

Sensitivity to Temperature Differences

Sensitivity to temperature differences consists of two intensive series, one running from thermally indifferent pressure through warmth and heat to thermally indifferent pain, and the other running from thermally indifferent pressure through cool and cold to thermally indifferent pain.

Sensitivity to warmth extends over some 12° C. of temperature, i. e., from slightly above 33° C., taken as normal skin temperature, to about

45° C. Sensitivity to cold extends over an approximately equal range from 33° C. to about 20° C. The only conditions *required* to distinguish these experiences from each other and from the thermally indifferent series of experiences (pressure-pain) are characteristic differences in the patterns for warmth and cold. Neither the differences nor their physiological conditions have been described. That the physiological conditions accompanying the characteristic differences for experiences of warmth and cold are functionally opposed is shown by the fact that such experiences cannot occur together in time and place although either may occur together with pressure or pain.

At extreme temperatures, experiences aroused by warm and cold stimuli are identical and neither retains any vestige of the distinguishing features of warmth and cold. With punctiform stimuli an observer can distinguish them from the prick of a needle only because of the greater pressure element in the latter, but under ordinary circumstances there are other sensory cues present so that one knows whether he is dealing with hot or cold objects (44).

There is, then, an intensive series of possible experiences ranging from slight pressure to severe pain. If an experience of intermediate intensity has a certain characteristic modification, it will be experienced as warmth, and if an opposing modification is present, it will be experienced as cold. It cannot have both in the same place at the same time, and without either it occurs as a thermally indifferent experience. It is doubtful whether experiences of warmth or cold ever occur without accompanying experience in the pressure-pain series (20, 53).

This hypothesis is supported by the facts brought out by experimentation with the "heat-grill" of Cutolo. The grill consists of a series of alternately warm and cold pipes, the temperatures of which are usually controlled by running water (16). Stimulation of an area of skin with the grill results in a pricky, pressury experience which at first produces a reaction like that to heat and experientially closely resembles heat except that it is not painful and an occasional warmth or cold temporarily displaces parts of the "heat" experience. The increments of experience necessary to mark warm and cold being functionally opposed, the close proximity of adequate stimuli for both prevents either pattern from being formed, and the experience continues without such modifications. Thus one experiences prick or bright points without the "tag" for cold, and a body of pressure without the "tag" for warmth. That the condition for setting up these increments is present, but held in balance, is shown by the fact that occasional shifts permit the introduction of momentary experiences of warmth and cold.

The older, qualitative theory accounts for the facts of the heat-grill in quite another way. The impressions of warmth and cold are mediated by separate receptors and fibers. Simultaneous stimulation of the two is accompanied by an experience of heat. This is possible under ordinary circumstances because a stimulus of high temperature arouses warmth responses and paradoxical cold responses, the two fusing to form an experience of heat[9] (13, 7, 6, 32).

A method for the solution of this problem lies in stimulation of the skin with punctiform stimuli over 45° C. If the qualitative theory is correct, reports should be in terms of warmth and cold, and, if the quantitative theory is correct, reports should be in terms of heat, prick or pain. This experiment has been performed once with results unequivocally supporting the latter theory (44). It appears from the results that application of stimuli of high temperatures is consistently accompanied by neither experiences of warmth nor of cold.

It is well known that experiences of warmth and cold may be aroused in fever, in emotional states such as fear or embarrassment, in goose-flesh, prickly heat, and under many other circumstances. The possibility that such experiences are accompanied by actual changes of temperature within the skin sufficient to account for their arousal is remote. It appears that any condition which is adequate to set up the requisite conditions within the tissue is adequate to the arousal of the correlated experience (42, 59).

The qualitative theory of warmth and cold has its roots deeply in the whole matter of specialized receptors. The assumed correlate for cold is Krause's end-bulb and that for warmth has varied from one organ to another. Extirpation experiments have always failed to discover any such relation, and Dallenbach has recently reported an exhaustive study of this matter, again by extirpation. His results are convincing (17). The tissue beneath large numbers of cold and warm spots was examined histologically and there was not a specialized ending of any kind found within the whole area of tissue taken from three observers.

The only nerve-endings found in the cornea of the eye are Krause's end-bulbs and free nerve-endings. This fact coupled with the report that the cornea of the eye is sensitive only to cold and pain led many to look for a relation between sensitivity to cold and Krause's end-bulb. Goldscheider, Fuchs, and others, however, showed that the sensitivity of this area is not limited to cold and pain and it has recently been shown that the cornea of the eye, if stimulated with a loose strand of cotton, is sensitive to contact and pressure (27, 44).

It appears from the evidence, as at present constituted, that experiences of warmth and cold are modifications only of general patterns of felt experience and that the physiological conditions correlated with these two experiences are functionally opposed. Temperature changes, as well as mechanical and electrical changes, are accompanied by the particular conditions within the tissues correlated with discharge of the receptors, and such a discharge will occur with changes in temperature whether or not the physiological conditions correlated with their characteristic differences (warmth and cold) are present.

---

[9] Von Frey reports that with warm, punctiform stimulation (about 45° C.), he would *occasionally* arouse an experience of cold (26, p. 172). "Paradoxical cold," as such an experience is called, was first described by Strümpell (51) and has been verified by Lehmann (34), Boring (10), and others. Von Frey states that, when he gets such reports, stimulation is usually not at a cold spot but near one.

PAIN

Despite the lack of much positive evidence, it is probably true that more time is required to arouse pain than to arouse indifferent pressure or contact experiences. Murray (39, 40) found this to be true, and Titchener (55, p. 152) states it as a fact. Adrian's records (2) show that a high frequency of impulses alone is not accompanied by painful experience but that the increased frequency must also be of greater duration, that is, the difference in the record between a (presumably) contact experience and a (presumably) painful experience is noticeably a difference in the duration of the increased frequency of impulses, the painful experience being of longer duration.

Both Titchener and Adrian attempt to account for these facts in terms of specificity of receptors. Titchener makes the statement that "the sensations proceeding from the pain organs are not necessarily painful" and that all pains do not "hurt" (55, p. 155). Adrian says that "it is conceivable that a momentary stimulation of the pain receptor may sometimes evoke a sensation which is not painful" (2, p. 49).

"Non-painful pain" is not attractive as an explanatory principle. Both authors suggest summation of impulses in some center as an explanation of pain, but such a suggestion does not compensate for assigning two qualities to a single fiber unless the evidence is positive upon the point. It is greater economy to drop specificity of receptors and supplement with some principle which depends upon the frequency and duration of impulses such as that included in the theory of chronaxie.

It is known that pain, at high intensities, is accompanied by emotion or is itself an emotion, the experience as a whole breaking up into analyzable parts consisting of the directly aroused experience, the reactions, the experiences aroused by such reactions, etc.

At all intensities it is known that pain is characteristically unpleasant. It seems possible, if not probable, that the unpleasantness itself consists of the experiences aroused by typical patterns of reflexly aroused responses. Unpleasantness is not easily analyzed out of the whole pattern of experience, does not easily fall apart as in the case of intensive emotional accompaniment. Under some circumstances unpleasantness is confused with pain, and the possible identity of "sensory" and "affective" pain has been the subject of a protracted argument in psychology.

At all intensities it is known that the direct, intensive experience which "is painful" or which "hurts" is accompanied by reflex activity in the tissues directly involved. The experience accompanying such reflex activity is not psychologically analyzable from the experience directly aroused.

We conclude that a sufficiently intensive experience becomes painful when it is of sufficient duration and when certain reflex activities are aroused. The "intensity of the pain" then varies directly with the intensity of the directly aroused experience and with the pattern of the impulses accompanying the reflex activities.

It has been stated that pain does not show the phenomenon of adaptation but this is true only in a very special sense (55, p. 154). It often happens, in the case of pain, that the tissues are injured so that the withdrawal of the stimulus affords no immediate relief. Conditions internal to the tissues may tend to increase the amount of stimulation, and the experience, as a whole, may become persistently painful. Yet adaptation becomes complete, even for pain aroused under such circumstances, long before the tissue heals although the "sore spot" may be more sensitive to pain than normal tissue until it is healed.

Localization and Other Learned Responses

Localization, posture, and orientation are learned, and such learning involves more than a single group of receptors. Usually both feeling and vision and often other senses are involved. Improvement in reaction to place, hence presumably learning, begins very early in infancy and continues, with modifications, throughout life. Localization is more easily investigated in adults by experiments upon relearning, and such studies have been made by Stratton (50) and Ewert (23). Genetic studies have been made but often upon children so far advanced that progress is difficult to measure (25, 46, 48).

The fact that one is not conscious of the processes which condition localization responses presents no difficulty as this aspect of learning is not disputed. Another example is the ability to distinguish the sound of one musical instrument from another without realization of the criterion. In positing cortical locus as the criterion for localization, we use the only neurological principle known which meets the conditions necessary for such learning.

The rôles played by different organs in these learned responses is not well understood. That the receptors of vision, feeling, and the non-acoustic labyrinths, hearing, and even smell are at times involved, and that memory or processes of conditioning play a large part, is generally accepted as fact, but the limits of the individual receptors in this respect are not known. Much less is it known to what extent the cutaneous, kinaesthetic, and visceral receptors furnish adequate criteria for localization independently of each other.

Localization by feeling is relatively accurate, less so than by vision but more so than by any other sense. Objects held in the teeth or in contact with a finger-nail or even with the end of a stick held in the hand are correctly localized. These mediating objects excite receptors within the tissues of the body, and the type of experience is recognized and is provocative of response as in learning. Localization systems based upon the activity of organs not used together do not correspond. An object held in the teeth may be accurately localized by the tongue, but this localization is a hindrance, rather than a help, if the finger is called upon to verify it. Such interference indicates two learned systems not in mutual relationship. Aristotle's illusion involving identification of the number of objects felt between two crossed fingers is evidence that the objectivity, which depends upon localization in this instance, is learned.

Kinaesthetic sensitivity is generally very accurately localized, especially in customary movements, while visceral sensitivity is relatively poorly localized. The known facts tend in general to support a theory of learning because new systems of activity are at first clumsily executed, and the most refined localization occurs in those parts most used, most aided by other organs, and when stimulation occurs in the usual manner. We do not see the skin of our back nor the viscera nor do we frequently use our ability to localize in those parts. Movement of the limbs, especially the hands, and localization upon the skin of the limbs and hands is constantly aided by vision, is much used, and is most accurate.

Discrimination of Two Points. If an area of the skin is successively stimulated by two points simultaneously applied, beginning together and at each trial being spread farther apart, there appears a phenomenon much like that of two spreading lights in vision. The two experiences are first felt as one and then separate slowly, taking distorted intermediate forms, such as that of a dumb-bell, before finally breaking apart to be felt as two discrete impressions. The distances at which the two points must be separated, to be experienced as two, vary upon different parts of the skin. The acuity of feeling in the skin ranges, according to Weber, from the volar side of the fingers, where such sensitivity is greatest, through the lips, palm, forehead, back of the hand, forearm and leg, upper arm, etc. The tongue is more sensitive than any of the skin areas named and the skin of the back is the least sensitive of any general area of the body surface (58).

In terms of the older qualitative theory it has been pointed out that the experience of discreteness must depend upon the intervention of at least one functional unit between the two stimulated areas. The functional unit, as used in this connection, is conceived to be a single nerve-ending or receptor. The experience of the intermediate shapes indicates that the separation is much greater than one intervening functional unit, and the evidence seems to show that not a single organ but a small area containing very many receptors is stimulated. Hence it seems more reasonable to suppose that two such areas are experienced as discrete when they are far enough removed to have no overlapping elements and that they approach unity as a matter of degree, not as a matter of steps, or, as stated under the qualitative theory, in a single step. In no case may it be assumed that by determining the "two-point limen" there is discovered the number of receptors involved.

SKIN SENSITIVITY

The types of experience aroused in the skin are varied and, unlike kinaesthetic and visceral experiences, are characteristically vivid, clear impressions. The sensitivity of the skin is not wholly independent of other parts, such feelings often being supplemented by those from the deeper tissues, and experiences of the deeper tissues are sometimes localized upon or referred to the skin as in the case of the "referred pains."

The skin may be affected as a whole or locally, that is, a pressure here, an area of cold there, etc. The same object may at one time feel cold to one part of the body and warm to another part as, where the two hands have been adapted to different temperatures, placing of both in water of a temperature between the two will feel warm to the chilled hand and cool to the warm one.

Active and passive touch are usually recognized as different in functional importance. Active touch is usually accomplished with the tips of the fingers or with an object such as a probe although other parts of the skin surface may be used. Passive touch may, of course, occur upon different parts of the body.

Many typical experiences have been analyzed into more simple collocations, e.g., Bentley (8) has shown that cold and pressure, in certain modes of occurrence, arouse a feeling typical of wetness, and Sullivan and Cobbey (52) have shown that warmth and pressure similarly may arouse an experience of oiliness. Zigler (60) has analyzed stickiness into collocations of pressure, and (with Meenes, 61) roughness and smoothness into other smaller complexes of pressure experiences. Tickle has been shown to be similar to contact experiences and to be aroused with complications of conditions of stimulation such as movement, double stimulation, etc. (43) This type of work is important in showing the manner in which patterns of experience are felt and the nature of the concepts built upon them.

Movement. The conditions for felt movement upon the skin are similar to those for vision, consisting of temporal and spatial relations between individual experiences. An object moving across the skin or a series of points stimulated in sequence arouse the perception of movement equally well, and such movement is generally of the type which in vision we speak of as "phi" movement. In the flashing type of movement, common in tickle, itch, pulsing pain, and similar experiences, we have a type of experienced movement comparable to "gamma" movement in vision.

Acuity. The acuity of the skin for form discrimination is markedly inferior to that of vision. In cases of mechanical stimulation this is due to the diffusion of stimulation occasioned by the pliable nature of the tissues involved and, in thermal and other types of stimulation, it is due to the intrinsic lack of form in the stimulus object or condition. Many studies upon the reliability of judgment of form, distance, and size, in filled and open spaces, have been reported.

Adaptation. Adaptation occurs locally and generally, and adaptation-time varies with the intensity and duration of stimulation. It is consequently shortest for warm, cool, and slight pressure experiences, and longer for heat, cold, and painful experiences. Adaptation-time for skin sensitivity is characteristically short as compared with kinaesthetic and visceral experiences. General or seasonal adaptation to temperature differences seems to occur in that if the temperature on a summer day drops to 50° F. the cold is felt as rather intense, but if it rises to that temperature after a cold snap in winter one feels delightfully warm. Adaptation to the pressure of clothing is often cited as an example of similar general adaptation to pressure, but in this case the adaptation seems to be, at least in part, a matter of attentive adaptation as one becomes conscious of such pressures whenever attention is directed to them. Adaptation to different barometric pressures, however, is general, e.g., altitude, and the principle here involved seems to be the same as for general adaptation to temperature differences.

The skin may be adapted to a temperature as low as 11° C. and to a temperature as high as 39° C., and such adaptation shifts the zero or indifference point. Temperatures higher than that of adaptation will feel warmer, and those below will feel cooler.

KINAESTHESIS

Kinaesthetic sensitivity consists for the most part of pressures, characteristically of low intensity and not vivid as are the experiences from the skin. Posture is maintained and learned movements are executed without being accompanied by much psychological experience from the muscles, tendons, joints, and bones themselves, attention following the visual experiences which usually accompany somatic activity. In the absence of vision, however, such learning may be acquired and maintained at a higher level of accuracy.

Experiences arising from these tissues include pressures, pains, aches, strains, and similar experiences, but warmth and cold and their combined forms such as smart and wet are not known, and it is probable that the conditions for setting up these experiences are not present in the deeper tissues. Thrills, shudders, tremors, and like experiences probably involve kinaesthetic sensibility to some extent, but how much of such experience may be an accompaniment of activity within the skin and how much is to be correlated with deeper sensitivity is not known.

The course of the nerves from these regions is not well known. It is believed by some that the sensory fibers extend from the muscles through nerves which also carry fibers from the skin, but the evidence for this appears to be incomplete. Most of the evidence tends to show that all sensory nerves from somatic sources run from the receptors to the cord, that in the cord they take a special pathway, independently of the nerves of the skin, to connections within the cord and to the higher centers. Study of the nerve tracts has not suggested the separation of these fibers according to the types of sensitivity either within the cord or within the peripheral nerves.

Position and Movement. Without other aids one is able to tell, with a fair degree of accuracy, the position of the limbs and any movement they may make. This is also true of the head, presumably, but being complicated by the sensitivity aroused with activity of the internal ear, it is difficult to tell how much is due to one and how much to the other. There is positive evidence which points to the joints as the most sensitive elements in the judgment for, if the muscles, tendons, and skin are slightly anaesthetized, the judgments are not greatly affected, but, if the joints are so treated, the judgments are much less accurate. It is probable that under ordinary circumstances all of the possible elements are operative but that the judgment depends more upon the impulses from the joints than upon those from other parts. If one moves the tongue without touching any part of the internal mouth, its position is accurately judged, showing that the articular sensibility is not necessary to the judgment. Movement of the limbs is more accurately estimated if there is resistance to the action.

Resistance and Weight. If we exert a force against gravity we lift a weight; if in any other direction we offer resistance. The accuracy of judging resistance to forces is greater for active than for passive resistance, and it makes some difference as to the manner in which the force is applied. In both cases the subject's judgment improves with practice.

All such judgments involve learning. There is no "sense of weight" but practice in estimating, especially if the errors are noted (usually by the aid of vision), makes it possible to judge accurately such position, movement, resistance, or weight.

Concepts of time and space are in part built upon the experiences of feeling. Extensity of felt experience and the form of patterns actually aroused are both experienced as spatial. Movement and position involve space and permit the building of a concept upon them. Vision has dominated in forming such concepts, however, and feeling has probably done little more than to modify them in a few details.

Claims for a "vibratory sense" have been made because the tissues of the body transmit vibrations which are felt but which cannot be accounted for by the combination of qualities of feeling as ordinarily understood. A special sense becomes unnecessary under the present theory, as the feelings are easily accounted for as physiological and psychological patterns.

Visceral Sensitivity

Visceral sensitivity is characterized by typical, rather elaborate patterns, such as those of hunger, thirst, nausea, colic, and others connected with the respiratory, digestive, circulatory, and reproductive systems. There are also isolated, less frequent, smaller complexes, and there are the feelings characteristic of certain diseased conditions. Parts of the viscera may be exposed to stimulation, which upon the skin would be effective, without arousing conscious experience. Some such parts may be cut, burned, pressed, and stretched without arousing any feeling and a few of these tissues, as far as known, are without sensitivity. Others, although apparently insensitive to the usual modes of stimulation, are extremely sensitive to their peculiar, adequate conditions of stimulation.

The general types of experience aroused with stimulation of the viscera are the same as for other parts of the body, consisting of experiences that are pressure-like or painful or combinations of the two. There is evidence that even warmth and cold are felt in some places. The nerves containing most of the fibers originating in the viscera are carried in the vagus. The massive movements are usually accompanied by long adaptation-times, and most of the feelings of the viscera are only vaguely and very inaccurately localized.

Affective Qualities. The affective experiences may be wholly or partly visceral in origin. At moderate intensities they are vaguely felt and poorly localized, but in greater intensities they are less difficult to observe. The difficulty in observation arises from the fact that such feelings are reflexly aroused and are conditioned to the perception of some object or state. When attention shifts from the pleasant object to the pleasantness for itself, the pleasantness is likely to disappear from consciousness with the pleasant object. At times, however, such experiences become more tangible as thrills and similar experiences and verge into emotional states.

Pleasantness, as a psychological experience, consists of discrete bright points of experience in the general nature of a thrill but usually much less intensive. It is vaguely localized in the upper part of the body and quickly adapts or fatigues. Unpleasantness is similar but characteristically duller, heavier, more of the pressure type of experience and is localized in the abdomen or in the lower part of the body (41).

The problem of mixed feelings remains unsolved. There is no apparent reason why elements of the two patterns should not be present at the same time. The available evidence tends to show that they never are, and such evidence sets up a presumption of functional opposition in the conditions requisite to the arousal of such patterns of experience.

REFERENCES

1. Adrian, E. D. The impulses produced by sensory nerve endings, Part 1. *J. Physiol.,* 1926, **61**, 49-72.
2. ————. The impulses produced by sensory nerve endings, Part 4. *J. Physiol.,* 1926-7, **62**, 33-51.
3. ————. The basis of sensation. New York: Norton, 1928. Pp. 122.
4. Adrian, E. D., & Zotterman, Y. The impulses produced by sensory nerve endings, Part 2. *J. Physiol.,* 1926, **61**, 151-171.
5. ————. The impulses produced by sensory nerve endings, Part 3. *J. Physiol.,* 1926, **61**, 465-483.
6. Alrutz, S. *Upsala läkaref. förhandl.,* 1897, **2**, 340-359; *Mind,* 1898, N. S., **7**, 141-144; *Skand. Arch. f. Physiol.,* 1900, **10**, 340-352.
7. Alston, J. H. The spatial condition of the fusion of warmth and cold in heat. *Amer. J. Psychol.,* 1920, **31**, 303-312.
8. Bentley, M. The synthetic experiment. *Amer. J. Psychol.,* 1900, **11**, 405-425.
9. Blix, M. Experimentelle Beiträge zur Lösung der Frage über die specifische Energie der Hautnerven. *Zsch. f. Biol.,* 1884, **20**, 141-156.
10. Boring, E. G. Sensations of the alimentary canal. *Amer. J. Psychol.,* 1915, **26**, 1-56.
11. ————. Cutaneous sensation after nerve-division. *Quar. J. Exper. Physiol.,* 1916, **10**, 1-95.
12. Botezat, E. Ueber Sinnesdrüsenzellen und die Funktion von Sinnesapparaten. *Anat. Anzeiger,* 1910, **37**, 513-530.

13. Burnett, N. C., & Dallenbach, K. M. The experience of heat. *Amer. J. Psychol.*, 1927, **38**, 418-431.

14. Cajal, R. Histologie du system nerveux, etc. (Trans. by L. Azoulay.) 2 vols. Paris: Maloine, 1909.

15. Carr, H. Head's theory of cutaneous sensitivity. *Psychol. Rev.*, 1916, **23**, 262-278.

16. Cutolo, F., Jr. A preliminary study of the psychology of heat. *Amer. J. Psychol.*, 1918, **29**, 442-448.

17. Dallenbach, K. M. The temperature spots and end-organs. *Amer. J. Psychol.*, 1927, **39**, 402-427.

18. ————. A bibliography of the attempts to identify the functional end-organs of cold and warmth. *Amer. J. Psychol.*, 1929, **41**, 344.

19. Dallenbach, K. M., & Burnett, N. C. The experience of heat. *Amer. J. Psychol.*, 1927, **38**, 418-431.

20. Dimmick, F. L. On the localization of pure warmth sensations. *Amer. J. Psychol.*, 1915, **26**, 142-150.

21. Donaldson, H. H. On the temperature sense. *Mind*, 1885, **10**, 399-416.

22. Erlanger, J., Bishop, G. H., & Gasser, H. S. Experimental analysis of the simple action potential wave in nerve by the cathode ray oscillograph. *Amer. J. Physiol.*, 1926, **78**, 537-573.

23. Ewert, P. H. A study of the effect of inverted retinal stimulation upon spatially coordinated behavior. *Genet. Psychol. Monog.*, 1930, **7**.

24. Forbes, A., Campbell, C. J., & Williams, H. B. Electrical records of afferent nerve impulses from muscular receptors. *Amer. J. Physiol.*, 1924, **69**, 283-303.

25. Franz, S. I. Accuracy of localization of touch stimuli on different bodily segments. *Psychol. Rev.*, 1913, **20**, 107-128.

26. Frey, M. von. Beiträge zur Sinnesphysiologie der Haut. *Ber. d. königl. sächs. Ges. d. Wiss., Leipzig, math-phys. Kl.*, 1895.

27. Fuchs, E. *Wiener. Med. Jahrb.*, **4**.

28. Goldscheider, A. Die specifische Energie der Temperaturnerven. *Monatssch. f. prac. Derm.*, 1884, **3**, (cf. *Ges. Abh.*, 1898, **1**, 53-76).

29. ————. Histologische Untersuchungen über die Endigungsweise der Hautsinnesnerven beim Menschen. *Arch. f. Anat. u. Physiol., Physiol. Abt., Suppl.-Bd.*, 1886, 191-227.

30. Haggqvist, G. Histophysiologische Studien über die Temperatursinne der Haut des Menschen. *Anat. Anz.*, 1913, **45**, 57.

31. Head, H., & Rivers, W. H. R. A human experiment in nerve division. *Brain*, 1908, **31**, 323-450.

32. Kiesow, F. Zur Analyse der Temperaturempfindung. *Zsch. f. Psychol.*, 1901, **26**, 231-246.

33. ————. The problem of the condition of arousal of the pure sensation of cutaneous pain. *J. Gen. Psychol.*, 1928, **1**, 199-212.

34. Lehmann, A. G. L. Die Hauptgesetze des menschlichen Gefühslebens, etc. (Trans. by F. Bendixen.) Leipzig: Reisland, 1892. Pp. 356.

35. Lotze, H. Gründzüge der Psychologie: Dictate aus den Vorlesungen. Leipzig, 1881. Also: Outlines of psychology: Dictated portions of the lectures. (Trans. by George T. Ladd.) Boston: Ginn, 1886.

36. McCouch, G. P., Forbes, A., & Rice, L. H. Afferent impulses from muscular receptors. *Amer. J. Physiol.*, 1928, **84**, 1-15.

37. Meyer, M. F. The fundamental laws of human behavior. Boston: Badger, 1911.

38. Müller, J. Handbuch der Physiologie des Menschen für Vorlesungen. 2 vols. Coblenz, 1830.

39. Murray, E. A qualitative analysis of tickling: its relation to cutaneous and organic sensation. *Amer. J. Psychol.*, 1908, **19**, 289-344.

40. ———. Organic sensation. *Amer. J. Psychol.*, 1909, **20**, 386-446.

41. Nafe, J. P. An experimental study of the affective qualities. *Amer. J. Psychol.*, 1924, **35**, 507-544.

42. ———. Psychology of felt experience. *Amer. J. Psychol.*, 1927, **39**, 367-389.

43. ———. Dermal sensitivity with special reference to the qualities of tickle and itch. *Ped. Sem.*, 1927, **34**, 14-28.

44. ———. A quantitative theory of feeling. *J. Gen. Psychol.*, 1929, 2.

45. Pendleton, C. R. The cold receptor. *Amer. J. Psychol.*, 1928, **40**, 353-371.

46. Peterson, J. Local signs as orientation tendencies. *Psychol. Rev.*, 1926, **33**, 218-236.

47. Ranson, S. W. The anatomy of the nervous system, etc. Philadelphia: Saunders, 1927. Pp. 425.

48. Renshaw, S. The errors of cutaneous localization and the effect of practice on the localizing movement in children and adults. To appear in *J. Gen. Psychol.*, 1930, **3**.

49. Stratton, G. M. Vision without inversion of the retinal image. *Psychol. Rev.*, 1897, **4**, 341-360; 463-481.

50. Strughold, H., & Karbe, M. Die Topgraphie des Kältesinnes in der Mundhöhle. *Zsch. f. Biol.*, 1925, **83**, 515-534.

51. Strümpel, —. *Deutsch. Arch. f. klin. Med.*, 1881, **28**.

52. Sullivan, A. H., & Cobbey, L. W. An experimental study of the perception of oiliness. *Amer. J. Psychol.*, 1922, **33**, 121-127.

53. Titchener, E. B. Experimental psychology. Vol. I. Qualitative experiments. Part II. Instructor's manual. New York: Macmillan, 1901. Pp. 456.

54. ———. The psychology of feeling and attention. New York: Macmillan, 1908. Pp. 404.

55. ———. A text-book of psychology. New York: Macmillan, 1919. Pp. 565.

56. ———. Notes from the psychological laboratory of Cornell University. *Amer. J. Psychol.*, 1920, **31**, 212-214.

57. Trotter, W., & Davies, H. M. Experimental studies in the innervation of the skin. *J. Physiol.*, 1909, **38**, 134-246.

58. Weber, E. H. Annot. anat., p. 149. Handwörterbuch der Physiologie. Vol. 3, Pt. 2, p. 524. *Ber. d. sachs. Ges. d. Wiss.*, 1852, 85. We used Hermann's Handbuch: Hermann, L. Handbuch der Physiologie, etc. Leipzig: Vogel, 1880. Vol. 3, Pt. 2, p. 379.

59. Wunderli, —. Experimentelle Beiträge zur Physiologie der Tastsinnes. (Diss.) *Molesch. Untersuch.*, 1860, 7, 393. We used O. Funke in: Hermann, L. Handbuch der Physiologie, etc. Leipzig: Vogel, 1880. Vol. 3, Pt. 2, pp. 321-324.

60. Zigler, M. J. An experimental study of the perception of stickiness. *Amer. J. Psychol.*, 1923, **34**, 73-84.

61. Zigler, M. J. & Meenes, M. An experimental study of the perception of roughness and smoothness. *Amer. J. Psychol.*, 1923, **34**, 542-549.

CHAPTER 10

SPACE AND THE NON-AUDITORY LABYRINTH

JAMES QUINTER HOLSOPPLE
Formerly of Yale University and of Western Reserve University

Man's knowledge of his position in space and his behavior which is contingent upon changes in spatial position are so taken for granted in ordinary experience and are so little understood in their reality that the psychological and physiological textbook chapters concerning them are for the most part merely footnotes. And even these seem to have been quoted, often without quotation marks, from the treatises of old anatomists who were so generous as to supply hypothetical functions for organs unknown to others than themselves. The *derivatives* of change in position, such as the variations in visual content, differences in sound localization, kinaesthetic data, and like phenomena, have been fairly well worked out. But knowledge of these derivatives presupposes in the first place that position in space can be known and can be held experimentally constant. Whether this presupposition is warranted is at present largely a matter for philosophical speculation. In fact, there are many who would hold that these "derivatives" are the basic data and "space" is actually the derivative. In experimental psychology, however, as in the other more or less natural sciences we may overlook such questions as these. We can marshal the "facts" of everyday experience, add to them our laboratory observations, and establish a useful working system.

First, we might make an array of the experiences which the ordinary layman would expect us to explain. The elevator, the railway train, the airplane, the merry-go-round, all furnish us with interesting data. All provide us with perceptions, some of which we recognize as "true," others we know are "illusory." All agree in changing our spatial position, either in translating us through one or more of the traditional three dimensions or in rotating us around one of our bodily axes. In every case the experience lies outside of that comprising the ancient five senses. In every case systematic, marked, and involuntary movements occur. And in every case the man whose vestibular mechanism is not in good psysiological condition finds that his experience is different from the normal.

When we ride in an elevator we know when we start and when we stop. But if we are riding in an express elevator and the floors are not marked, it is exceedingly difficult to estimate one's position at any given moment. The same phenomenon is apparent in riding on a railway train. In this case one's uncertainty has been used to clarify the theory of relativity for those who will not understand it anyway, with the result that neither relativity nor vestibular psychology has profited. In the cases of the elevator and the train, after the effects of visual, auditory, cutaneous, kinaesthetic, and visceral stimulation *directly* due to movement have been accounted for, little remains to be explained by vestibular activity. This seems to be true for rectilinear movement in general so far as man is concerned. In the cases of the airplane, merry-go-round, and the like, the situation seems to be different.

For convenience we may consider the body as having three axes; one passing downward from the top of the head, one passing through the body from left to right and one from in front backwards. In the case of the toe-dancer, for example, the body would rotate around the first. The body would rotate around the second in exercises on a horizontal bar. Rotation around the axis last mentioned rarely occurs in ordinary behavior, except for movements of very slight excursion. Rotations around one or another of these axes occur frequently; they produce marked effects on behavior and on "consciousness," and these effects cannot be explained without reference to the vestibular mechanism. One's peculiar experiences in an airplane and on the merry-go-round are instances in common life of these rotational phenomena.

The tragic predicament of the flier, upside down in the fog; the pathetic state of the syphilitic whose VIIIth nerve involvement prevents comfortable locomotion; and the much overworked but quite legitimate story of the lost traveler in the blinding snow or trackless forest; all of these, together with the common experiences mentioned above, impress upon the experimental psychologist that more is demanded of his system of space and the vestibule than he can find in his predecessor's textbook.

If we accept the "stimulus-response" type of psychological analysis, the problems of space and the vestibule are as clear as the present data are inadequate. With respect to normal, adult, human psychology a representative list might be stated as follows.

1) Is there a perception of space which cannot be explained in terms of sensory data without recourse to the vestibule?

2) Do perceptual data which involve the vestibule all bear upon space perception, or are they in any respect independent of perception of space?

3) What, in detail, are the characteristics of the perceptions which follow vestibular stimulation?

4) What are the features in the environment which stimulate the receptor mechanism?

5) What is the receptor mechanism and how does it work?

6) What are the responses, other than perceptual, which follow vestibular stimulation?

7) What are the necessary conditions for the modification of the vestibular response?

Without attempting to answer these questions adequately we can at least take them in order and indicate an experimental psychological approach to the solution of each.

There *is* a perception of space which cannot be explained without recourse to the vestibule. The student should verify this fact for himself, and be careful to follow the procedure *exactly* as given. If he does as instructed, he will add greatly to his perceptual experience. If he modifies this procedure, he may subject himself to serious discomfort. First, he should seat himself in a revolving chair and obtain a helper who is *not* obsessed with the "practical joker" type of humor. Then blindfolded he should let himself be turned through 10 complete revolutions. The rate of rotation should be about one turn in two seconds. The turning should be as uniform as possible, without jerking. The stop should be not too jerky, but should not require more than one turn for complete deceleration. *Under no circumstances should the head be moved while rotating, nor should the head be moved until all perception of movement is gone.* This procedure should again be followed with the following changes. The body should be bent forward with the forehead resting securely upon the knees and the center of gravity well back in the chair. *Only five* turns should be given in this position. Immediately after stopping the chair, the person who is rotating it should indicate to the one being rotated that he should sit up. *Upon sitting up the eyes must remain closed and stay closed until all perception of movement is gone,* otherwise visceral discomfort may ensue. This same procedure may now be followed with the head upon the knees and at the same time turned at right angles, so that the head is facing in the direction of rotation. The foregoing precautions must be rigidly observed here also. During the Great War, when rotation tests for aviators were in vogue, a great deal of nausea occurred. This almost invariably followed the neglect of the precautions given above. If the student has followed these instructions as given and has observed carefully the data of movement and position, he sees that there is spatial experience which is contingent upon movement of the *head* as such. Further evidence will show that the observed effects are dependent upon *vestibular* stimulation localized within the cranial cavity. He is now aware, also, that at least part of this spatial experience does not fall within any of the categories of sensory experience *other* than the peculiar one which he may now label *vestibular.* Perhaps no detail of psychology illustrates better than this one the relative inadequacy of textbook psychology as compared with learning by experiment.

Even though these new experiences have been labeled for convenience "vestibular," the student should note that they may be only indirectly due to vestibular stimulation. This problem will be considered in a later paragraph.

The second question, whether or not all vestibular data pertain to space perception, is less easily answered. Nor will it be answered until more elaborate and painstaking experiments have been performed than have been reported so far. The difficulties lie in the fact that the perceptual pattern following vestibular stimulation is so complex, so many modes of sensory data are involved, and the analysis is so difficult, that a naïve experimenter does not know in his first attempts just what part of the experience is vestibular. The situation is analogous to a study of vision if the apparatus used were of necessity noisy, odorous, and painful.

If we keep in mind the fact that many unconscious bodily changes occur when the vestibule is stimulated, and the further fact that these changes may be perceived as such, it is clear that any analysis of the directly and purely vestibular component in the whole complex may be quite erroneous. Many adjectives have been used to describe the experience. The expressions "dizziness," "vertigo," "whirling," "sinking," "head-pressure," and the like show how varied is the pattern and how complex the experience. For the present we may assume that whenever the vestibule is stimulated there occur, in our experience, facts which do not seem to have other spatial reference than that of being in general inside our bodies. In this respect vestibular data are much like our "feelings" and our "emotions." In fact, the anatomical and physiological considerations involved do not preclude at all our treatment of vestibular phenomena as "feelings" rather than in the traditional class of "perceptions." On the basis of these facts we must accept the hypothesis that vestibular experience, although dependent in normal life upon spatial factors for its inception, and although described almost wholly in spatial terms, is in some particulars as independent of space *as such* as any other variety of experience.

The third question, concerning the details of the perceptions which follow vestibular stimulation, is particularly interesting. Were this chapter written about vision, or olfaction, or audition, this paragraph would constitute the major part of the discussion. In connection with vision, for example, we would point out, even though we could not define, the nature of brightness and of hue. We would distinguish between acuity and sensitivity. We would show how all the details of quality, quantity, intensity, duration, position in space, and position in time, combine in infinite variety to make up our visual world. Or, if we felt that these details are merely incidental to the pattern, we might show how they are but the artifacts resulting from our analysis. But in any case our discussion would deal with the same general set of facts, and the facts are much the same in kind whether we discuss vision, audition, or what not. If we peruse almost any textbook in general psychology, we find that this holds true only until we reach the vestibule. Here the content changes suddenly. No longer are these details significant. No longer do we speak of thresholds. Weber's law and Fechner's formula seem to have lost their significance. But more disastrous than any of these deficiencies is the fact that our psychological system has broken down. The responsibility for repair rests squarely upon experimental psychology. At present if we are honest we must admit that we lack not only data but even ingenious hypotheses. To be sure the term "threshold" is to be found in the periodical literature concerning the vestibule, but even here, as we shall show subsequently, it does not have a clear meaning with respect to the vestibular stimulus, and none at all so far as the perceptual data are concerned.

When we come to the fourth question, which concerns the nature of vestibular stimuli, we have reached more solid ground. Stimuli, as such, lie quite as much within the province of the physicist and physiologist as within that of the psychologist. However, one would hardly question the psychological necessity for understanding the environmental changes which give rise to different mental processes. The vestibular mechanism, like other receptor organs, may be stimulated in several ways. For the normal human, the vestibular mechanism responds to an angular acceleration around any of the bodily axes. It may respond to rectilinear movement, but this point has not been settled. It *certainly* responds to accelerations which occur in rotational movements. A great deal of confusion in our experimental literature is due to an apparent lack in understanding this point.

Rotation in itself does not constitute a stimulus for the vestibule. If one is rotated in a smoothly moving chair, and if he waits until the effects of his initial acceleration have worn off, and if the velocity remains constant, then *no vestibular data* are *perceptible* so long as the head is not moved. Nor do any unpleasant effects result from rotation under these conditions even when the turning is kept up for several minutes. Even now these points are generally disregarded in the experimental literature. Many authors seem to believe that it is the rotation or angular velocity which is important. The facts which are responsible for the persistence of this misapprehension may be summed up briefly as follows. Rotation in a clockwise direction is usually perceived as different from rotation in a counterclockwise direction, regardless of the axis around which rotation occurs. The perceptual data following rotation are sometimes different for different directions of rotation. The bodily changes following a cessation of rotation are, under prevailing conditions of experiment, peculiar to the direction in which rotation has occurred. The effects of rotation are, within certain limits, dependent upon the duration of rotation and upon the speed of rotation. The effects of frequently repeated rotation in one direction have been shown in some experiments to be different from the effects of rotation in the opposite direction. If these facts were the only ones at our disposal, rotation could legitimately be considered as the vestibular stimulus. At this point, however, we must consider rotation more analytically and add certain other facts to those just listed.

Any rotation to which a normal adult human would be subjected can be divided into three components. At present these components are *not* usually controlled separately in experimental procedure. They should be. First there is a period of angular acceleration, let us say in a clockwise direction. Then there is a period of rotation at a uniform velocity. Finally there is a period of deceleration. Of course, in any rotation there may be many accelerations and retardations, but these three are of necessity always present.

Let us suppose now that whenever a person is rotated in a single plane he receives two sets of stimuli, one upon acceleration and another of opposite kind upon retardation. Then if all of the facts summed up previously can still be accounted for, and if certain additional observations are explained which seem inexplicable when *rotation* is considered as the vestibular stimulus, we must agree that *angular acceleration* provides the stimulus.

The first fact to be considered is that rotation in a clockwise direction is perceived as different from one in a counterclockwise direction. Strictly speaking, this is true only when there is a single acceleration and a single retardation. If one is rotated so that he starts to move in a given direction and is then subjected to slowing and speeding up in the rotation rate, it becomes impossible for him to tell in what direction he really is moving. If we now suppose acceleration to be the stimulus we see that we distinguish rotations to the right from those to the left by virtue of the fact that we *start* to the right or left and not because we are moving in either of those directions.

The bodily changes which follow rotation have been observed to be different after one has been rotated to the right from those which follow rotation to the left. These will be discussed in greater detail in a subsequent paragraph, but for the present we may merely note that they comprise changes in tonicity of antagonistic muscles, one of a pair contracting while the other relaxes. For example the *rectus internus* of the right eye may contract while the *rectus externus* relaxes, thus causing the eyeball to turn inward. The same reasoning used in the just-preceding paragraph applies here also. If one is being rotated to the right and is slowed in his movements, the bodily changes are appropriate to the direction of acceleration and are not dependent upon the direction in which he is actually moving. The only reason for experimenters' having supposed that the post-rotation effects of turning in one direction are different according to the direction of rotation is that in their usual procedure when a subject *slowed down* in his movements he usually *stopped*. No matter how rapidly one has been turning nor how long, if the *retardation* is too gradual to serve as a stimulus, *no* different effects will be noted from rotations in opposite directions.

The third fact, that the effects of rotation are dependent in part upon the speed and duration of rotation, does not show that rotation is the real stimulus. It merely follows from the fact that in order to reach a high speed one must pass through a long or rapid acceleration, and that in stopping from a rapid turning one receives more effect from his retardation. With respect to duration and its effects the situation is a little more complicated. If the acceleration stimulus is followed very soon by a retardation, and if these stimuli are opposite in their operation, then the retardation would be expected to be less noticeable than if it occurred so long after the acceleration that the effects of the latter had completely disappeared. This is exactly what happens in actual experiments. Long rotations do seem to show more immediate effects than do short ones, but the lengthening is effective only up to a certain point. After this point is reached no further differences in the after-effects are noted regardless of differences in duration of rotation.

During the early days of the Great War when nystagmus tests were in use as measures for sorting out aviators, the idea that practice in rotation might effect the bodily changes subsequent to the rotation was unthinkable outside of a few limited psychological circles. At that time and for a number of years thereafter experiments were carried out in order to determine whether practice in rotation in only one direction would produce effects different from those resulting from practice in the opposite direction of rotation. Results of these experiments tended to show that it would. But if we assume that for rotation in a single plane there are really only two types of stimuli, a clockwise and a counterclockwise acceleration; also that both of these stimuli occur regardless of the direction of rotation, except for their being reversed in temporal sequence; also that the practice effects are due to factors coincident with the repetition of the stimulus; then we are upon the horns of a real dilemma. The solution of the problem hinges upon the fact that in the practice series the time relations of the acceleration and retardation stimuli effect the results of practice. In this connection the author has shown that it is possible by varying these time relations to produce practice effects which were thought to be due to practice in a given direction by actually rotating the subject in the opposite direction. Hence even this last reason for supposing rotation to be the vestibular stimulus has been shown to be due to inadequate observation and acceleration can be accepted in its place.

It is hardly necessary, in discussing vision or audition, to elaborate on the nature of the physical stimuli involved. The measures of intensity of *light* are obviously different from the measures of *brightness*. The intensity of sound as measured in terms of the wave amplitude in the vibrating medium is obviously independent to a certain extent of the psychological loudness of the sound. The student of psychology may be presumed to know his physics as a basis for psychological knowledge. But to psychological investigators of vestibular phenomena the nature of angular acceleration as a physical fact seems often not to have been clear. Velocity is distance traveled per unit of time. For rectilinear movement that would be miles per hour, feet per minute, and like expressions. Acceleration is the rate at which velocity changes. For linear acceleration if one travels at the rate of four miles per hour and during the course of the next succeeding hours increases his speed gradually until after three hours he is traveling at the velocity of nineteen miles per hour, he is said to have accelerated at the rate of five miles per hour per hour. It will be noticed that one cannot know anything about acceleration unless he knows at least two velocities and the time elapsing between them. And even under these conditions he will know only the *average* acceleration. These same considerations hold also for angular accelerations and velocities. In this case the unit of distance is the degree. Thus if one makes a complete revolution in two seconds, he will have had an average velocity of 180° per second. If it requires five seconds to attain such a velocity that the next complete revolution requires two seconds, then the average acceleration will have been 36° per second per second.

Unfortunately for the ease with which we can make psychological measurements, average values for a stimulus are seldom useful. The average light intensity upon the retina would be of little value in any determination in research upon vision in spite of the fact that uniform light sources are obtained fairly readily. Average acceleration means little for vestibular research at present because we have yet to construct the apparatus for laboratory use in which we can be certain that we will accelerate the subject uniformly. It is important for psychology that we determine threshold values for vestibular stimuli because it seems now that these values are the ones whose variations are responsible for many individual differences in practical life situations such as balancing, maintaining one's orientation, and the like. In neurological examinations, where nystagmic phenomena are becoming increasingly important for diagnosis, threshold determinations would be particularly useful.

After the student has really understood that there is such a component in his space perception as the vestibular, after he has made at least an attempt at its identification, and after he is thoroughly familiar with the environmental changes which are responsible for its existence, then, and preferably not before then, he should examine the receptor mechanism upon which the stimuli impinge. In many texts the discussions of the vestibule both start and finish with physiological deductions which have been made from anatomical observations. Valuable as both the deductions and observations are intrinsically, they are often misleading when carried far into the field of psychological explanation.

At this point it should be emphasized that no verbal description is an adequate substitute for physiological experiment, nor can any diagram or photograph substitute for a first-hand dissection. Diagrams of the vestibular and semicircular canal mechanism do not give one a clear picture of either the size or spatial relations involved. Large composition models of the ear or head are generally misleading also. Hence the student should look for an opportunity to visit a museum where there is available a preserved dissection of these organs. In this way he can understand some of the experimental difficulties which are due to the size of these receptor mechanisms. In many anatomical museums and libraries the student can find stereoscopic slides which show the relations of the semicircular canals and the vestibule to adjacent structures and to each other. If these museum specimens or slides are available, the following description should help the student to understand them. If they are *not* available, he will save time and have a better understanding of what he knows by omitting to read this description and passing directly to the discussion of the nonperceptual effects of vestibular stimulation.

The internal ear is usually called the labyrinth because of its complex shape and numerous cavities. In the internal ear are the endings of the VIIIth nerve. This nerve carries only afferent impulses, which are derived from stimulations to the organs of hearing and stimulations to the receptors of the semicircular canals and those of the vestibule proper. That part of the labyrinth which contains the cochlea we are not concerned with here. Nor are we concerned with that part of the VIIIth nerve which transmits impulses connected with audition.

The labyrinth is conventionally considered as consisting of two parts anatomically, the osseous labyrinth and the membranous labyrinth. The osseous labyrinth is merely a group of connected cavities which lie in the petrous part of the temporal bone. The temporal bone is situated at the side and the base of the skull. Its upper and anterior part is thin and translucent, but that part which is wedged between the occipital bone in the rear of the skull and the sphenoid in front is pyramidal in shape and contains the labyrinth. This portion is called the *pars petrosa* or *pars pyramis.* Within this osseous labyrinth is a series of connected sacs and ducts which are called, when taken together, the membranous labyrinth. The cavities are lined with periosteum, and separating the membranous labyrinth from the bone is a fluid called the perilymph. The sacs and ducts of the membranous canal are filled with a fluid which is called endolymph.

The labyrinth is conventionally divided into three divisions spatially. Each of these divisions contains its bony and membranous portions. They are the cochlea, the vestibule, and the semicircular canals. Only the last two concern us in this chapter.

The vestibule is ovoid in shape, the longest axis being about 5 mm. and the shortest about 3 mm. long. The membranous sac within the vestibule is divided into two parts, the utricle and the saccule. The utricle is the larger of the two and into it the semicircular canals open. Hair-cells line the utricular and saccular membranous structures and upon these hair cells lie small masses of grains of calcium carbonate held together by a gelatinous mass. The endings of the VIIIth nerve which supply the vestibule are in contact with these hair-cells. In the lateral side of the vestibule is an opening which is normally kept closed by the base of the stapes and annular ligament.

The bony semicircular canals are above and behind the vestibule and open into it. They are unequal in length, the longest being the posterior canal, which measures 22 mm. The shortest is the horizontal, which is about 15 mm. long in most cases. The other canal, the superior, is generally about 20 mm. long. The bony diameter measures about eight-tenths of a millimeter, and within this lies the membranous canal in each case. The lumen of the membranous canal is about two-tenths of a millimeter. These measurements are of considerable importance when one considers the theories of their functioning to which we shall refer later. Each canal has two openings into the vestibule, except that the superior and posterior canal have one duct in common. These canals lie in general in planes at right angles to each other. At one end of each canal, just before it opens into the vestibule, there is a thickening called an ampulla. In these ampullae are distributed the nerve-endings to the canals.

The theory of functioning of this mechanism which is most current in the neurological textbooks has long since been discarded by more careful experimenters. But because of its prevalence a caution against it is necessary. This theory is that, when rotation occurs, the endolymph flows through the canals, thereby stimulating the nerve-endings in the ampullae. In the 1927 edition (3rd ed.) of one of the most widely used clinical neurology texts we find the following statement: "Any movement of the endolymph in relation to the semicircular canals or vice versa is called lymphokinesis, and constitutes the irritation which stimulates the semicircular canal or canals in which lymphokinesis takes place." According to this theory the 20- or 30-second duration of nystagmus following rotation is supposed to be due to the continuation of this "lymphokinesis." We shall point out later the possibility of stimulating the canals in other ways than by rotation, but for the present we may merely note that cold-water irrigation acts as a stimulus. According to this same authority, "The endolymph in this part will sink down, and thus cause lymphokinesis according to the same thermosyphon principle which is utilized in water-cooling on light motor-cars." The unvarnished fact is that for all the anatomists have told us about this receptor system we are still in the dark about how it works.

We must now attempt to answer our sixth question, which concerns those effects of stimulation which are not directly perceptual. The fact that stimulation of the labyrinth produces changes in various parts of the body is psychologically important for two reasons. In the first place, these bodily changes are themselves perceived and constitute important items of mental content. In the second place, they serve as indices of vestibular action.

It should be noted at this point, in view of the presumption that the student is somewhat familiar with the receptor system, that the term "vestibular" as used in this chapter is a convenient abbreviation for "vestibular and semicircular canal." Labyrinthine is not a good term because the labyrinth includes the cochlea. Furthermore, in spite of the excellent physiological work which has been done on lower animals, we are not clear as to the difference between the human functions of the vestibule and the canals.

As previously mentioned, the main bodily effects of stimulation to the vestibule are changes in tonicity of antagonistic muscles. For practical purposes we may confine ourselves to a discussion of two sets of such changes, namely, nystagmus and past-pointing.

Nystagmus is the series of eye movements which follow an angular acceleration. If one seats himself in a revolving chair and turns himself slowly through about ten or fifteen degrees, if he closes his eyes and rests his fingers lightly upon the lids, he will feel the eyeballs rotating in their sockets in the direction opposite to that in which he is being turned. Were one teleologically inclined, this movement would be easily explained by supposing that this movement is due to the necessity of keeping the visual stimulus in an unchanged retinal position. Certainly it is true that many discomforts would ensue upon the absence of this movement. In walking, in driving an automobile, even while sitting quietly in a chair, there are many head movements occurring involuntarily. The slightest movement would result in a blurred visual pattern if the eyes did not keep their fixation relatively independent of head movements.

This movement of the eyeballs in the direction opposite to that of the rotation proceeds at a rate which is determined by the rate of acceleration. But the extent of movement is obviously limited by the physical structure of the eye. When the ball has reached the limit of its excursion, the opposite movement occurs. This return of the ball to its normal position is independent of vestibular excitation. It is a rapid movement, and, as in all quick eye movements, vision is inhibited while the movement occurs. If rotation is continued, these alternate drifting and jerking movements constitute the series of movements which we call nystagmus.

There are three kinds of nystagmus which follow vestibular stimulation and which are obviously different even to the casual observer. There is first the horizontal nystagmus, which is a drifting and jerking of the eyes to the right or left. This is the form which is usually studied because the stimulation for it results in a minimum of unpleasant accessory effects. The vertical and rotary nystagmus movements are directed as their names imply, and are easily distinguished from horizontal nystagmus and from each other.

The nystagmus which occurs after an acceleration at the beginning of a rotation is the same as that which occurs at the slowing down except that the directions of the slow and quick movements are reversed.

The commonest measure of nystagmus is its duration. Barany deserves the credit for popularizing this measure. He found that if a normal subject is rotated at the rate of about a turn in two seconds with acceleration rates unknown for both starting and stopping, one may expect the post-rotation nystagmus to last anywhere from 16 to 32 seconds. Shorter or longer times indicated a pathological nervous condition in his opinion. Duration, however, is not the only variable. The frequency of the individual movements varies to some extent independently of the duration. The amplitude of the movement also varies from individual to individual and for the same individual from time to time.

Past-pointing, as well as nystagmus, has been extensively studied in connection with vestibular stimulation. If one extends his arm directly in front of the body, blindfolded, and lowers his arm quickly he can touch his knee with fair accuracy. If, however, he has just been rotated he will miss the point at which he aims by several inches. This is called past-pointing. The effect gradually wears off after a rotation so that the accuracy of aim will increase even though the subject is unaware that he has made any errors.

In connection with nystagmus and past-pointing there is another phenomenon which deserves consideration. This is the illusion of rotation. This might be considered either with the perceptual effects of vestibular stimulation or with the non-perceptual effects because we do not know now what the real stimulus for it is. Everyone remembers his childhood experiences of whirling himself around until equilibrium is lost and the subsequent view of the world in motion. It is clear how this phenomenon occurs. During the drift movement of the eyes we see the world moving by. During the return movement we do not see it come back because the movement is a quick one and we are temporarily blind. But even with our eyes closed we feel ourselves rotating. We could explain this feeling or illusion in terms of kinaesthetic sensitivity from the eye movements were it not for the fact that the illusion persists long after the nystagmus has disappeared.

Concerning the musculature other than the eye and arm we need say only that for the whole body similar changes seem to occur. They are not all so easily discernible as the two mentioned, but any attempt at locomotion or balancing shows how far-reaching the effects are.

The viscera, as noted above, are also affected by vestibular stimulation. Nausea is easily induced, especially by stimulation of the vertical and posterior canals. A slight but very general and persistent discomfort is sometimes felt for many hours after stimulation. These effects, however, are easily avoided with careful experimental technique.

The last question propounded at the beginning of the chapter, concerning the conditions necessary for the modification of the vestibular response, is important in view of the fact that this response has been supposed to have permanence and stability far in excess of those experimentally established. So much work has shown that nystagmus time as well as other reactions of rotation change markedly under practice conditions that one need do no more now than mention the fact. In respect to practice effects vestibular responses have much the same status in our knowledge as do other so-called reflexes. We know, for example, for those who have recently been rotated the nystagmus time is less than for those who have not. We know, too, that the decrease is dependent in part upon the number of rotation periods which have been experienced. General physical conditions affect the reductions due to practice, but the nature of the conditions and the amount of their effects remain to be determined. The practice reductions have been shown to be most marked within a relatively short period after practice and seem gradually to disappear after practice has been discontinued. In general, the various other responses to vestibular stimulation show the same susceptibility to modifying conditions which are present during the period of stimulation as does nystagmus.

To this point we have concerned ourselves with only a few aspects of our general problem. We have tried to orient ourselves with respect to the place of labyrinthine (vestibular and semi-circular canal) sensitivity in the general field of psychological data. We have tried to place it on a psychological basis for consideration instead of simply adding it to our psychological system as an anatomical or physiological appendage. In order to avoid confusion we have considered only the normal adult human individual. But there are many facts which throw light on the vestibular mechanism and which are not essentially normal, human, or adult. A discussion of the more important of these follows.

Phylogenetically, the development of the vestibular apparatus is extremely interesting. The complexity of movements of equilibration and orientation does not increase uniformly as the evolutionary series develops. But as the demands which are put upon the organism in maintaining its spatial position change, so we notice changes occurring in the mechanism of the labyrinth. Changes in the softer tissues of the body are obvious and frequent, but skeletal similarities show relationships between animals which seem to differ widely in the zoo, and skeletal changes do not occur with great fluency. Hence a study of the bony labyrinth shows a phylogenetic diversity of structure which is almost startling. A few hours spent with stereoscopic slides from the library of a comparative anatomist will not be wasted for the psychologist if he will take the trouble to familiarize himself with the habits of locomotion of the animals whose labyrinths he views.

Not only do the structures of the bony labyrinth change as one goes through the evolutionary series, but there are some corresponding variations in the development of the nervous structures as well. Too much emphasis must not be placed upon the quantity of nervous tissue used in the distribution of impulses from any particular set of receptors. In general, as a special sense becomes less important in the life of a species the proportion of central nervous tissue used in distributing and organizing its afferent impulses diminishes. For the vestibular receptors the central connections are found in Dieters' nucleus and in the triangular nucleus of Schwalbe. If one compares the area of these portions of the central nervous system in the lower and the higher primates, it does appear in general that those of the former are a little more extensively developed, especially when their life is largely arboreal. The large area devoted to these central connections of the vestibule is quite noticeable in the case of the lemur. The lemur is the lowest of the primates. It is a small animal, lives in Madagascar, and is arboreal for most of the time. It leaps from one tree to another and has a constant need for nicety of balance. Also the proportion in the brain of Schwalbe's and Dieters' nuclei is high. It should be noted here that although there is an evolutionary progression in the vestibular areas the differences are far less marked than is the case in many other brain areas. But then it is also true that man, for all his sedentary habits, is still a balancing animal and in constant need of precision in his equilibration and his orientation.

Owing to the fact that the quantity of experimental work on the vestibule of animals has been so much greater than that done upon man, and in general the quality so much higher, there is a great temptation to write one's vestibular human psychology in terms of animal experimentation. The difficulty lies for the most part in the fact that the functions seem not always to be identical. In many animals the receptor organs are larger, not only relatively so to their body size but actually so. The skull is thinner and gives readier access for operative procedure. The response mechanisms are more readily studied. They can be rotated almost indefinitely. And, perhaps the greatest advantage of all, they give no conflicting verbal reports of their experiences. In the case of animals—iron filings have been substituted for the calcium carbonate of the otolith organs in the vestibule proper; the whole labyrinth has been extirpated; the individual canals have been stimulated directly by pressure upon one or another of them; whole generations of rats have been raised in rotating cages; the separate muscles of the eye have been detached from the ball and their contractions measured directly and independently of the movement of other muscle groups; in fact, the only limitation upon animal experimentation has been that imposed by the lack of ingenuity of experimenters.

The development of vestibular responses from birth to adult life is a field which is much in need of experimental cultivation. It is certainly true that vestibular reflexes in early life are much less marked than after the adult stage is reached. Children whirl themselves violently about, bring nearly all their canals into a position susceptible for stimulation, and are merely amused by effects which would not be tolerable to an adult. This can hardly be due solely to stronger viscera but must be a result of a less thoroughly functioning labyrinth. "Age norms" for nearly every other mental and physical process have been at least tentatively accepted, but so far no one has determined a "mental age" for vestibular responses. This is all the more surprising in view of the obviously low Pearsonian or Spearmanic correlation coefficient which this function would show when compared with intelligence quotients and like series of values. One explanation for this deficiency might be found in the impossibility of reducing vestibular measurements to pencil and paper tests and in the requirement upon the measurer that he know something about the methods and techniques of experimental psychology.

A complete understanding of vestibular problems requires that one should be familiar with contributions which have been made by physicians whose interests have been mainly in pathological vestibules. To say that nystagmus studies, if properly carried out, serve as a complete index to the condition of the central nervous system, is probably something of an exaggeration. It is clear, however, that the functions of the vestibule in so many persons are so little affected by the ordinary conditions of their lives, and the ramifications of the efferent paths from the vestibule are so numerous, that any neurological study is aided by a knowledge of whatever peculiarities may be shown in vestibular reactions. Abnormal nystagmus phenomena certainly aid in the localization of brain tumors. This fact is due not only to the involvement of vestibular areas themselves but also the nystagmic abnormalities which occur when adjacent lesions affect vestibular functions.

The inadequate or unusual stimuli for the vestibule play a large part in medical diagnosis based on vestibular responses. All receptor organs are susceptible to stimulation by other forces than those which normally initiate their action. These forces are usually called inadequate stimuli. In the case of vision, for example, a sudden blow upon the eyeball will cause one to see a flash of light. In the case of the vestibule and semicircular canals the responses may be elicited not only by an angular acceleration, but also by irrigating the ear with hot or cold water, or by passing an electric current through the head, or whenever there is a fistula into the labyrinth the vestibular receptors may be stimulated by inflation of the external canal of the ear. These methods of stimulation are sometimes employed clinically in addition to the rotation.

The irrigation effects are called "caloric" since the nature of the response is dependent upon the temperature of the water which is used for the irrigation. If one is lying upon his back during the irrigation and then suddenly sits erect, the effects are not the same as those produced by rotation with the head down and then sitting erect after the rotation has ended. The exact nature of the caloric reaction is not known, but the theory or any derivative of the theory of lymphokinesis fails sadly as an explanation. It is very difficult to regulate the stimulus applied in this way and the visceral effects are liable to be considerably more disturbing than in the case of rotation.

When one passes an electric current through the head, the stimulus is called "galvanic." There are many objections to the use of this method by one who is unfamiliar with electrical apparatus and the vestibular mechanism. When this method is used, one electrode is placed upon the tragus of one ear and the other is put upon the opposite side. It is interesting to note that the faradic current will not serve as a stimulus. Only a few milliamperes are required to elicit a response. The major objection to this method, so far as theoretical considerations are concerned, will be obvious to the experimenter if he will act as subject himself and allow the use of a little too much current. In this case he will notice, when the current is turned on, a number of interesting phenomena. He will see a flash of light, smell a pungent odor, taste a salty something, hear a clicking sound, and feel a "jolt" of his head, all simultaneously. Obviously he has received an inadequate stimulus to all of his cranial senses. There is no reason to suppose that something of the same kind does not occur even when he has received just enough current to stimulate his vestibular mechanism without producing perceptible effects on the other sensory mechanisms. No one knows where the current goes as it passes through the head after leaving the positive electrode. All that is known in this case is that when one reverses the positions of the positive and negative electrodes the nature of the vestibular responses seem also to be reversed.

The method of stimulating the labyrinth by direct pressure of air through a fistula is unimportant psychologically since for practical purposes one can never depend upon getting pathological cases of the right sort in the laboratory.

There are many points of technique and method in vestibular research which are of little consequence to the general reader but which should be carefully considered by the experimental student whether he wishes to do vestibular work himself or merely desires to be able critically and intelligently to evaluate the experiments of others. It is assumed that he is familiar with the general status of knowledge concerning the vestibule of the normal human adult and furthermore that he is not wholly ignorant of the bearing upon the knowledge of the sorts of cognate information which we have indicated. The more important of these points of technique and method are outlined in the following paragraphs.

The selection of a working hypothesis is of primary importance. It is obvious from the most cursory examination of many reports of vestibular work that the experimenter had no clear idea of the point which he wished to settle. If the problem concerns the effect of angular acceleration as compared with linear acceleration, then it seems clear that the problem should not be complicated by introducing unknown relationships between clockwise and counterclockwise accelerations. So far as it is possible, every variable other than the one in which the experimenter is interested should be held constant. In vestibular work, as in every other type of experiment, a few hours of preliminary thought may save one from many subsequent hours of doubt. One hour spent in devising a method of holding one variable constant while another is the object of experiment may save one the trouble of devising statistical methods of allowance which will themselves require time and in the end leave the results open to question. The type of problem upon which one works is relatively unimportant. The important point is keeping the problem simple. So far as experimental psychology is concerned, the most glaring deficiencies lie in our ignorance of normal human adults.

The selection of subjects raises some real questions. Whether they should be practiced or not before the experiment when the experiment does not purport to settle a question of practice effects is one typical problem. In general, it is true that the naïve subject will find the whole situation so new that many paradoxical reactions may occur. On the other hand the use of long-practiced subjects implies that the experimental results will not yield valid inferences regarding the responses of subjects who are unpracticed. The sheer physical limitations upon the supply of subjects may settle this point for the experimenter. In case they do, the experimenter should keep the alternatives in mind and should not "rationalize" in explaining why he chose his subjects as he did.

The apparatus to be used naturally depends upon the question to be settled. One should, however, not overlook the fact that most experiments have been performed by graduate students and that the question to be settled more often than not depends upon the apparatus which is available. If the subject is human and the stimulus is acceleration, the devices at the disposal of the experimenter are limited. The ordinary piano-stool revolves, but as a laboratory instrument here it is useless. The otologists have used a chair which is not satisfactory for experimental purposes, chiefly because it is hand-driven and because the subject is not firmly supported. The acceleration cannot be controlled in a hand-driven chair, nor can one maintain a uniform velocity of rotation after the desired acceleration stimulus has been administered. Unless the head is provided with a more satisfactory support than that used in any other type of laboratory work, head movements are sure to occur in spite of the subject's best intentions. In order to eliminate these difficulties Dunlap has devised an electrically driven chair and attached a head-rest to the chair with steel braces.

The Dunlap chair is mounted upon a shaft to which is attached a steel plate parallel with the seat. As the chair turns, this plate rotates between two sets of electromagnets. One set of these magnets rotates constantly, the other set is stationary. As an electric current is passed through the moving magnet, the chair is rotated. In order to stop the chair the circuit is broken through the moving magnets and made through the stationary set. By increasing the speed of the motor the speed of the chair can be regulated, and by varying the current through the clutch and brake magnets the positive acceleration in starting and the negative acceleration in stopping can be varied within certain limits. This apparatus was a great improvement upon the old Barany chair but its use is subject to some criticism. The plate which rotates between the clutch and the brake magnets must have a slight clearance, and this clearance is just enough to cause a perceptible sway and consequent variation in the rotation rate if the chair does not rest upon a precise level. There is a constant source of annoyance in the location of the motor which cannot be placed in the axis of rotation. Since it is slightly off center, it is always possible for the subject to estimate his rotations by reacting to differences in sound intensity. The contact between the clutch plate and the magnets must be metallic, for a certain amount of slipping is necessary if one is to avoid too sudden acceleration. But this slipping of the metallic surfaces is dependent upon the nature of the surfaces which naturally varies with the amount of use they have had, whether a drop of oil from the machinery has fallen upon them, and other like factors. Hence the absolute amount of acceleration is never quite definitely known. These disadvantages have been discussed in detail because this apparatus, for the purposes for which it was intended, is the best one now available.

Dodge has devised an ingenious chair which makes possible a study of responses to less intense accelerations. This chair lacks many of the disadvantages of the Dunlap chair but possesses others which are quite as unfortunate. The Dodge chair is driven by a small electric fan which is attached to the chair and acts as a propeller. The obvious difficulties of controlling the acceleration rates are not eliminated by this type of propulsion, nor does one have great confidence in the uniform speed of small motors as they are operated under the usual laboratory conditions.

In all vestibular work the control of non-vestibular stimuli is of the utmost importance. This is apparent in the light of our doubts about which responses are directly due to vestibular stimulation and which are due to other factors. Unless eye movements are to be noted, the eyes should be well shielded. Noises of any kind will provide clues to one's position and will elicit those responses of bodily orientation to sound which always occur when the sound is not clearly localized. Any bodily movement during rotation changes the action of centrifugal forces upon the flexible musculature and arouses kinaesthetic data in which one is not interested but which may alter the data which are presumed to be strictly vestibular.

After all of one's problems concerning hypotheses, subjects, and apparatus have been tentatively solved, there still remain the intricate problems of recording and measuring. Heretofore the most reliable vestibular data have been those involving nystagmus, but measurements even of these are susceptible to grave error.

Nystagmus, as such, is easily observed and identified by the most untrained. Its measurement is another matter. As we noted above, the frequency of nystagmic movements decreases as the series draws to a close. The observer who is timing the duration, stop-watch in hand, must decide which of the jerks of the eye is the last one. There is no distinguishing feature of this last movement save that it is the last. If these jerks come a couple of seconds apart toward the end, it is obvious that the error of measurement will be high. This difficulty can be avoided, but at the cost of considerable effort, by timing the duration of the nystagmus by recording each of the final movements upon a kymographic record. Another difficulty in nystagmic observation inheres in the fact that the amplitude of the movement decreases with practice in rotation. The use of lenses which prevent fixation tends to keep the nystagmic movements constant with respect to excursion but they also change the criteria of the experimenter's observations. These difficulties have led to attempts at photographic registration which have met with indifferent success. Dodge invented a device in which small mirrors rested upon the closed eyelids. This shows the movements of the eye under the lid but it also shows movements of the lid, and the apparatus naturally cannot distinguish between lid and ball movements. If the eyes are photographed while open, a high illumination is required. This inhibits the normal nystagmus and introduces discomfort for the subject.

The sources of error in method and technique which have been outlined above do not begin to exhaust the possibilities. They merely serve to indicate the directions in which one must look while outlining an experiment for himself or while interpreting the results of another investigator. On the other hand, one must not suppose that because there are so many difficulties therefore no vestibular work is reliable. One must merely maintain an attitude which is not too credulous while he encourages advances in our knowledge.

For the student who wishes to work in a psychological field where the problems have not resolved themselves into petty discussions of minute points the vestibule offers major questions yet unanswered. If one wishes the widest possible opportunity for exercising his ingenuity in devising apparatus and techniques, the vestibule and semicircular canals offer unlimited possibilities. If one's concerns are philosophical and space is interesting, he can find in the vestibule an empirical basis which he dare not overlook. If applications of knowledge to human conduct seem important, we have in the vestibular responses a set of behavior patterns which are involved in every situation calling for equilibrium or orientation. To be sure, these advantages in vestibular work are not without their cost. It is true in general that where knowledge is deficient it is usually hard to obtain. If the problems were easily answered, they would have been answered long ago. The writer after having spent many months upon one vestibular problem reached the conclusion that nystagmus from a negative acceleration, when it occurred soon after an acceleration in the opposite direction, was inhibited, and in proportion as it was thus inhibited it failed to show a diminution such as previous experiments indicated it ought to show when it had been through a practice series. One may imagine how disheartening was the experience of being told then by the head of his department that his discovery was about equivalent to finding that people who have not been rotated but who use their heads normally in turning this way and that do not suffer a decrease in their nystagmus times by so doing. Thus many experiments upon the vestibule have been elaborations upon the obvious, and so many more will be. But on the whole the chances are no worse for the vestibule than elsewhere. In view of our lack of knowledge the rewards for genuine vestibular discovery should be high.

In writing this chapter the attempt has been to provide a basis for understanding the problems of the vestibule rather than to compile a handbook or encyclopedia of information. The references at the end indicate where details may be found.

REFERENCES

For a summary of work published prior to 1921 see:

1. Griffith, C. R. An historical summary of vestibular equilibrium. *Univ. Ill. Bull.*, 1922, **20**, No. 5. ✓

From 1921 to 1926 the following works are recommended:

2. Dodge, R. A mirror recorder for photographing the compensatory movements of closed eyes. *J. Exper. Psychol.*, 1921, **4**, 165-174.
3. ————. Habitation to rotation. *J. Exper. Psychol.*, 1923, **6**, 1-35.
4. ————. Thresholds of rotation. *J. Exper. Psychol.*, 1923, **6**, 107-137.
5. Dunlap, K. A rotator for vestibular and organic stimulation. *J. Comp. Psychol.*, 1921, **1**, 365-367.
6. ————. Adaptation of nystagmus to repeated caloric stimulation in rabbits. *J. Comp. Psychol.*, 1925, **5**, 485-493.
7. Griffith, C. R. Are permanent disturbances of equilibrium inherited? *Science*, 1922, **56**, 676-678.
8. Holsopple, J. Q. Some effects of duration and direction of rotation on post-rotation nystagmus. *J. Comp. Psychol.*, 1923, **3**, 85-100.
9. ————. Factors affecting the duration of post-rotation nystagmus. *J. Comp. Psychol.*, 1923, **3**, 283-304.
10. ————. An explanation for the unequal reductions in post-rotation nystagmus following rotation practice in only one direction. *J. Comp. Psychol.*, 1924, **4**, 185-193.
11. ————. The importance of the analysis of the complete vestibular response. *Trans. Amer. Otol. Soc.*, 1925.
12. Laurens, H., & Miles, A. L. The effect of unilateral labyrinthectomy on rotational nystagmus. *Amer. J. Physiol.*, 1927, **80**, 150-156.
13. Maxwell, S. S. Equilibrium and labyrinth. Philadelphia: Lippincott, 1923. Pp. 163.
14. Pike, F. H. The function of the vestibular apparatus. *Physiol. Revs.*, 1923, **3**, 209-239.

15. Travis, R. C., & Dodge, R. Experimental analysis of the sensorimotor consequences of passive oscillation, rotary and rectilinear. *Psychol. Monog.*, 1928, **38**, No. 3.

16. Wilson, J. G. Physiology of the labyrinth. *Arch. Otolar.*, 1925, **1**, 231-240.

The literature since 1926 is excellently indexed and annotated in the *Psychological Abstracts,* published by the American Psychological Association and edited by Walter S. Hunter, Clark University, Worcester, Mass. See index headings, "Labyrinth," "Vestibule," "Semicircular canals," etc.

CHAPTER 11

HUNGER AND THIRST

W. B. Cannon
Harvard Medical School

In the functioning of our bodies most of the essential processes are managed automatically. For example, after the act of swallowing, the entire complicated process of digestion is thus controlled. The constancy of body temperature, the constancy of the acid-alkali balance in the blood, the constancy of blood-sugar and numerous other conditions which are of prime importance for natural existence, are assured by agencies which operate correctively whenever there is significant deviation from the normal state. Among these conditions is adequate provision of food and water. Food is continuously required in our bodies as a source of energy for muscular work and as a means of repair and renewal of worn or injured body structure. Water likewise is of the utmost importance; it forms the bulk of the blood and lymph and of the fluid inside of cells, it serves as a carrier for the transfer of food from the alimentary tract into the blood stream, it helps to make a lubricant between moving surfaces, and it plays an essential part in the control of body temperature. Clearly both food and water are among the first necessities of the organism. At all times, however, they are both being spent. Volatile waste from the burning of food material in the body is carried away by every breath. Other, non-volatile waste is always passing out through the kidneys. Serving as a vehicle for the discharge of this waste is water. Also water is continuously being evaporated from the respiratory surfaces, and by insensible perspiration from the skin as well.

Since food and water are steadily being lost from the body the only way in which a constant supply can be maintained is by means of storage and gradual release. Food is stored in the well-known forms of fat and body starch or glycogen, and probably also as protein in small masses in the liver cells. Water is stored in tissue spaces and in tissue cells. As need arises, these stored reserves are set free for use. The reserves themselves, however, must be replenished. It is the function of hunger and thirst as automatic stimuli to make certain that the reserves of food and water are maintained.

HUNGER AND THIRST VERSUS APPETITE

In any discussion of hunger and thirst it is important at the outset, as Cannon (3, chap. XV) has emphasized, to distinguish carefully between these sensations and appetite. The experiences of hunger and thirst are so well known that their definition may be deferred until later. Appetite has been regarded as a mild form of hunger or thirst. It belongs, however, to a quite different category. Appetite is dependent upon previous experience which was of such agreeable character that there is a desire for repetition of it. Thus, we have an appetite for a particular food or drink in accordance with the pleasure which it has given, although there is no real state of hunger or thirst. We eat a delectable dessert at the end of a long meal when we are not at all hungry, and we take a drink that gives us pleasure when we are not thirsty. The appetite for food and drink may lead to the taking of both nutriment and fluid far beyond bodily requirements. These appetites are in the nature of invitations to indulgence, and, as will be seen, are quite different from either hunger or thirst. If, however, an appetite does not lead to the taking of food or drink, then the much more exacting demands of hunger or thirst appear and act as impelling agencies which insistently require the eating of food or the drinking of water or watery fluids.

The Nature and Basis of Hunger

Hunger has been described, and most persons recognize the correctness of the description, as a very disagreeable ache or pang or sense of gnawing or pressure which is referred to the epigastrium, the region just below the tip of the breast bone. In a large group of persons there are almost invariably two or three who declare that they have never felt this sensation; possibly they have never been in a situation where a serious deficiency of food prevailed, or appetite, habit, and immediate provision have led to the taking of food before hunger has arisen. Experience of hunger, however, is so commonly recognized in the terms above mentioned that it may be so defined.

Theories of Hunger. The older theories of hunger assumed that it was a "general sensation" which was based on a general bodily need for food. It was supposed that the lack of food in the circulating blood directly stimulated the cells of the brain, among others, and that the sensation thus arose. According to this view, the common reference of the sensation to the vicinity of the stomach was explained as due to an association developed by the disappearance of the hunger pang when food was taken into the stomach. There are, however, serious objections to this view, which may be summarized as follows:

1) Careful studies of the blood have shown that there are no changes in it which are associated with actual bodily need. The content of sugar, fat, and protein in the blood stream remains within the common range of variation even at times when hunger may be prominent.

2) The cells of the brain are relatively insensitive to chemical stimulation; indeed, they are insensitive to most forms of stimulation except deprivation of oxygen. The brain may be cut and manipulated in a conscious individual without any experience of discomfort. Even if there were food deficiencies in the blood, it seems quite unlikely that they would affect the cells of the cerebral hemispheres.

3) In certain times of great general bodily need, as, for example, during fever, when food is not relished, when it is commonly not taken in con-

siderable amounts and may be ill digested, when there is actual wasting of the body—when, in short, hunger as a general sensation should be most intense—it is in fact usually absent.

4) The swallowing of indigestible materials, such as scraps of leather, bits of moss, or even clay, has been reported as a means of stopping the sensation of hunger. Since these do not provide food which would supply exigent brain cells, banishment of the sensation by them indicates that the sensation is not dependent upon general bodily requirements of nutrition expressed particularly in the requirements of the cerebral cortex.

In addition to the foregoing reasons for not crediting the theory that hunger is a general sensation, it is noteworthy that that theory does not account for the quick onset of hunger. Many persons have reported that the hunger pang or ache appears abruptly. That there could be any critical change in a general bodily state which would account for this sudden appearance of the phenomenon is highly improbable. Furthermore, by careful attention to hunger pangs, they may be observed to recur rhythmically; they come and go with a fair degree of regularity. There are no such variations in bodily needs. It is incredible that plenty of food material would be available for use in the body one minute, and a minute later a serious deficiency would occur, and perhaps a minute thereafter the supply would be adequate again. Opposed to the theory that hunger has a diffuse origin in a general bodily need, the quick onset and the periodicity of the sensation raise obstacles which cannot be satisfactorily set aside.

The theory that hunger is a general sensation, moreover, fails to explain the local reference to the epigastrium. The supporters of that theory, to be sure, have declared that it is fallacious to ascribe the pang to a change there because it may be a "referred sensation," like the referred sensation of tingling, for example, experienced by a man who has lost the part of his body in which the tingling seems to originate. The justification for this criticism, however, depends upon whether there is, in fact, a local change in the vicinity of the epigastrium that is temporally associated with the sensation and that might reasonably account for the hunger pang which is actually felt. There is, indeed, evidence that changes occur in the stomach which would account for the sensation.

The Local Origin of Hunger Pangs. Various theories have been offered to account for the origin of the hunger sensation by events in the stomach. The emptiness of the organ, the presence of hydrochloric acid within it, and the turgescence of the secreting cells in the gastric wall have all been suggested as possible causes of the pangs. It is obvious, however, quite apart from other reasons which might be cited, that the quick onset and the frequently recurring periods of hunger are not consistent with these views. The stomach would have to be empty and full within two minutes, hydrochloric acid would have to be present or absent within the same period, and the gastric glands would have to be turgid and flaccid with similar speed. All these suppositions involve impossibly rapid changes.

The observations of fairly frequent and rhythmic recurrence of hunger pangs, together with associated sounds of moving air in the stomach region,

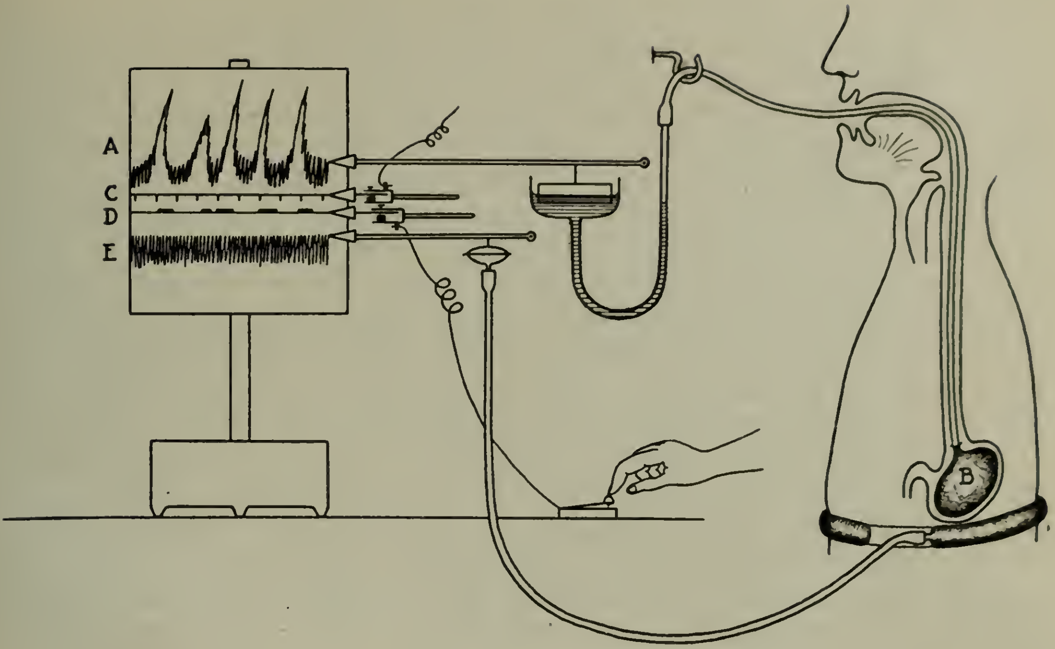


FIGURE 1

Diagram Showing the Method Used to Record the Gastric Hunger Contractions

A, kymograph record of the increase and decrease of volume of the gastric balloon, *B*. *C,* time record in minutes. *D,* record of the subjective experience of hunger pangs. *E,* record of the pneumograph placed about the waist; this record proves that the hunger contractions do not result from action of the muscles of the abdominal wall.

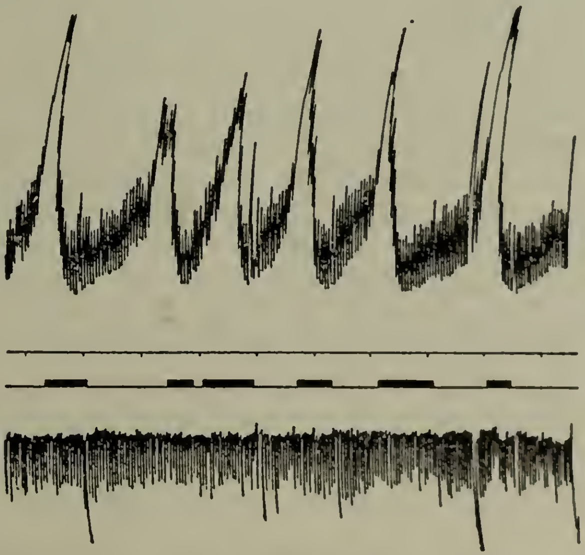


FIGURE 2

Photograph of the First Record of Hunger Contractions Associated with Hunger Pangs, Taken May 17, 1911

One-half original size. (From: Cannon, W. B. A consideration of the nature of hunger. *Pop. Sci. Mo.*, 1912, **81**, 304.)

led Cannon (2, p. 204) to the idea that the pangs might be due to strong periodic contractions of the muscle of the gastric wall. In 1911 he made observations on Washburn (4), who had accustomed himself to the presence of a rubber balloon in the stomach and to a tube in the esophagus, which led from the gastric balloon to a recording apparatus (see Figure 1). A pneumograph about the abdomen, which recorded abdominal motions, gave assurance that the pressure changes recorded in the stomach were not due to contractions of the abdominal musculature. A key in the subject's right hand was pressed when he felt the sensation of hunger. Volume changes in the gastric balloon, a time-marker for minutes, the signal which registered the sensation, and respiratory changes in the abdominal pneumograph were all recorded in the same vertical line on a moving kymograph drum. In Figure 2 is reproduced the first record obtained by use of this method.

As shown in Figure 2, there are powerful periodical contractions of the empty stomach, lasting approximately 30 seconds and recurring at intervals which vary from 30 to 90 seconds, with an average of about 60 seconds. As the reader will observe, the testimony of the subject that he experienced a hunger pang was not recorded until the contraction had nearly reached its peak. The sensation therefore was not the cause of the contraction—the contraction was the cause of the sensation.

The observations made by Cannon and Washburn were soon confirmed by Carlson (7). He studied the phenomena of hunger in a man with a gastric fistula, and also in himself by becoming accustomed to a gastric balloon as Washburn had done. In a series of interesting researches on human beings and on various kinds of lower animals, Carlson (6) and his students have brought out many new aspects of the hunger pangs and the relations of gastric contractions to them. Among the characteristics of the phenomena, they have shown that hunger usually begins with occasional weak contractions of the empty stomach, and that these contractions become gradually more vigorous and appear at shorter intervals until an acme of activity is reached which may end in a true spasm of the gastric muscle. Either the single contractions or the spasm may be associated with the typical, unpleasant ache or pang or gnawing sensation which has long been recognized as the experience of hunger. After the acme of activity has been reached, the stomach usually relaxes and remains inactive for a period, whereupon it starts again with occasional weak contractions, and the cycle which has just been described is repeated. Carlson's subject with a gastric fistula reported that distention of the stomach by means of the balloon or rubbing the mucous lining with a smooth object did not cause sensations of hunger unless these procedures produced contractions. It is definitely the contraction of the gastric muscle that brings about the hunger pang.

By observations on the pressure changes in the stomach as recorded from an intragastric balloon, at the same time that the gastric contractions were examined by means of the X-rays, Rogers and Martin (10) have found that there are two types of activities in the stomach during the hunger pang. At the height of a pang the lower portion of the stomach may be so strongly contracted that the cavity there may be completely obliterated. In addition, there may be unusually powerful peristalsis at such a rate that a second wave appears before the first one has disappeared.

Carlson and his students have studied with care the conditions which might influence the hunger contractions. The contractions, they have found, occur in a typical manner during sleep—a fact which explains in part the occasional restlessness of sleep. They are stopped by chewing and temporarily by swallowing, as Cannon (2, p. 204) also had observed. It is well known that strong emotional states abolish all digestive activities of the alimentary tract; these states likewise abolish the gastric contractions associated with hunger. The contractions are weakened and may be completely checked by smoking, and apparently in this respect the efficacy of smoking varies according to the "strength" of the tobacco. The direct introduction of alcoholic beverages—undiluted beer and wines, and 10 per cent alcohol—into the stomach through a gastric fistula caused cessation of the hunger contractions and also lessening of the tonus of the empty stomach. Very vigorous muscular exercise inhibits the periodic waves of hunger, but after the inhibition they are likely to recur with an intensity greater than before. Moreover, strong tightening of an abdominal belt nearly always leads to a stoppage of the contractions, when they are of weak or moderate strength, that lasts from five to fifteen minutes. The stoppage may be partial or complete, but in either case the contractions reappear despite continued pressure of the belt.

The interesting problem arises as to what induces the stomach, while empty, to contract with a vigor which is greater than that observed in the ordinary peristalsis recurring regularly during the digestion of a large meal. Reverting to the view that hunger is an expression of a general bodily need of nutriment, Müller (9, pp. 524-530) has expressed the opinion that this need affects a part of the brain, which thereupon induces the gastric contractions and thus indirectly causes hunger pangs. That the powerful contractions are not due to extrinsic nervous influences, however, has been proved by completely isolating the stomach from the brain and spinal cord and later observing that the typical contractions occur, though at somewhat longer intervals than in normal animals (5). Again, it is known that the elective source of energy for muscular contraction is carbohydrate food—glycogen or sugar. It seemed possible that a deficiency of this energy-yielding material might be signalized by excessive contractions of the smooth muscle of the stomach. Bulatao and Carlson (1) observed that if the sugar concentration of the blood was reduced about 25 per cent by use of insulin, the hunger contractions became more intense—an observation which has been noted subjectively in human beings who have received an overdose of insulin. On the injection of sugar into the blood stream the hunger contractions were abolished. Although this evidence is suggestive, it cannot be regarded as definite proof that the hunger contractions are due to lack of available sugar, for the amount of sugar used to abolish the contraction was excessive both in amount and concentration as compared with normal blood sugar. Furthermore, if an accessory stomach is made by isolating a part of that organ from its main body, the two stomachs may exhibit complete independence with regard to hunger contractions. Vigorous activity may be present in one during quiescence in the other (6). This points to a local automatism as the primary factor rather than to a condition of the blood, for the character of the blood flowing to the main stomach and the accessory stomach is necessarily the same. The problem as to the cause of the contractions, therefore, still remains unsettled.

The conception of the local, peripheral origin of hunger, developed in the foregoing paragraphs, has been criticized by Hoelzel (8), who has emphasized a "central" origin of the sensation. He has not made clear, however, his definition of hunger; he intimates, indeed, that it is not a sensation but an urge to action, and he furthermore presses the point, which has been recognized by various observers, that hunger contractions may exist without producing the characteristic sensation. It is true that hunger does urge to action—indeed, as a rule, every disagreeable sensation is attended by an impulse which leads to such action as would result in abolishing the unpleasant experience. To magnify the impulse so extensively as to have it comprise all the subjective phenomena of hunger would appear to be a mistake. Again, in arguing that hunger contractions are not the source of hunger pangs because they may exist without producing the characteristic sensation, Hoelzel fails to recognize that many circumstances may prevent stimuli from having their usual effect. A ticking clock may not produce the sensation of ticking because other stimuli are simultaneously affecting the nervous system or because, for instance, the nervous system may be dulled by sleep. That the ticking is not heard at times does not prove that it does not produce the sensation when it is heard. There is good evidence that by paying attention to hunger contractions and watching for them, their subjective results in hunger pangs may be intensified. It is also certain that by appeals to other interests hunger sensations may be largely eliminated from consciousness.

The Nature and Basis of Thirst

Thirst is a sensation referred to the inner surface of the mouth and throat, especially to the root of the tongue and to the back part of the palate—a highly unpleasant sensation of dryness or stickiness. The tongue cleaves to the teeth or to the roof of the mouth, or a lump seems to be present at the back of the throat, and there is repeated swallowing in order to dislodge it. All accounts of well-marked thirst agree that its main characteristic is the dry mouth. King, who described the experience of a United States cavalry troop which was lost for more than three days on a desert in the southeastern part of the United States, reported that on the third day there was no secretion of fluid into the mouth, that attempts to chew resulted in food sticking in the teeth, and that sugar did not dissolve on the tongue.

Other evidence which relates local dryness of the mouth with thirst is found in experiences which result in a rapid evaporation of the moisture of the mouth or a diminished production of fluid there. For instance, the breathing of hot, dry air, as well as prolonged speaking or singing, or the chewing of dessicated food such as dry crackers, will lead to the sensation of thirst and the desire for something to drink. Fear and anxiety likewise are associated with dryness of the buccal mucous membrane and may cause distressing thirst. Besides these local conditions, however, there are certain general bodily states which may produce the sensation. Profuse sweating, for example, or the excessive loss of fluid from the body in the course of disease as in the diarrhea of cholera or in the abundant discharge of water from the kidneys in diabetes, will provoke thirst to an intense degree. It is noteworthy that the withdrawal of milk from the body by a nursing infant induces in the mother a dominating desire for water. And after extensive hemorrhage thirst is tormenting—in the hospital wards after a battle the cry for water is almost universal.

In accordance with the evidence that thirst is due to local dryness of the mouth and also accompanies states of water need in the body as a whole, there have developed theories concerning the nature of the sensation similar to those developed for hunger; some experimenters have advocated the view that thirst is of local, peripheral origin, and others have argued that it is a general sensation.

Theories of Thirst. The theory that thirst has a general or diffuse origin has been based on the supposition that when there is a diminution of the water content of the body all parts may suffer from the deficiency; supporters of this theory explain the local reference to the back of the mouth as due to an association between the feeling of dryness there and the fairly quick disappearance of the feeling when water is taken. The analogy between this view and that of a conditioned local reference of hunger to the stomach is obvious. The evidence supporting the idea that thirst has a diffuse origin is in the main derived from certain general modes of treatment of thirst which affect the body as a whole and which abolish the sensation. For example, thirst may be promptly banished by the injection of water under the skin or into the intestines. The introduction of water by these routes does not moisten the pharynx, and yet the desire for water disappears. In commenting on this evidence the point may be made that the treatment, though general, may change the local conditions in the mouth and throat so that the sensation no longer arises from that region. If this should, indeed, be true, the abolition of thirst by general treatment would not prove that thirst has a diffuse, general origin, but rather would prove that it has a local origin in the pharynx. The main fact which must be kept in mind is that the thirsty man does not complain of a vague general condition, he complains of a parched and burning throat. That phenomenon must be accounted for in any theory of thirst which may be offered.

Müller (16, pp. 530-537) has suggested that lack of water increases the concentration of crystalloid substances in the blood, that this excites a region in the base of the brain, and that this in turn discharges impulses which cause contractions of the esophagus which give rise to thirst sensation. Records of esophageal contractions are reported as being more pronounced after prolonged deprivation of water than they are when the water supply of the body is abundant. These records show, however, marked oscillations of the intensity of esophageal activity—such oscillations as cause the *hunger* sensation to be felt as recurrent and rhythmic. Since thirst is not experienced in this manner, but is a steady and persistent discomfort, the evidence that esophageal contractions cause the sensation seems unreliable. Furthermore, Müller's interpretation fails to take into account the central fact of thirst, which has been emphasized by all observers —the distressingly dry mouth and pharynx—and also the empirical testimony that this dryness of the mouth is, in fact, the source of the sensation. For example, persons suffering from severe thirst because of great losses of water through the kidneys have had their distress relieved when the sensitiveness of the nerve-endings at the back of the mouth was destroyed by painting the region with cocaine. Moreover, sipping a small amount of water and temporarily moving it about in the mouth will stop the sensation. Also, holding in the mouth a substance which causes secretion of saliva—a bit of lemon, for instance—will lessen thirst. None of these procedures supply water to the body, and yet the distress is mitigated. Their efficiency in bringing relief does not suggest, however, any rational account of the relation between the local dryness and the water lack in the organism as a whole. Such an account must explain how the dry mouth as a local condition may be a means of automatically indicating bodily need and of automatically leading to the satisfaction of the need. Clearly, an arrangement must be sought which, when the body requires water, would cause the mouth to become dry. This arrangement might reasonably be expected to be found only in animals which are continuously and rapidly losing water and which therefore must have repeated renewal of the water supply in order to maintain a normal condition. These clues have been followed (12, chap. XVI, or 13).

Among vertebrates the fishes alone are strictly water-inhabiting animals; amphibia may live in air but they must remain near water so they may resort to it if they tend to become dry; only reptiles, birds, and mammals are strictly inhabitants of the air. The water in which fishes live wets constantly their body surface, and furthermore it is constantly passing through the mouth and out through the gills in providing for respiration. This stream of water moving through the mouth offers a continuous supply, and under such circumstances it seems improbable that thirst would arise. The conditions are quite different in the air-inhabiting forms. The body surface, which is in contact with the air, is dry. No longer is there a water current which keeps the mouth moist. Instead, there has been added to the structures of the head a nasal chamber through which an air current passes. It is noteworthy that at the back of the mouth this current crosses the pharynx to reach the windpipe or trachea (see Figure 3). Microscopic examination shows that the lining of the nasal chamber and of the trachea is well provided with cells which secrete a mucous fluid. On the other hand, the lining of the back of the

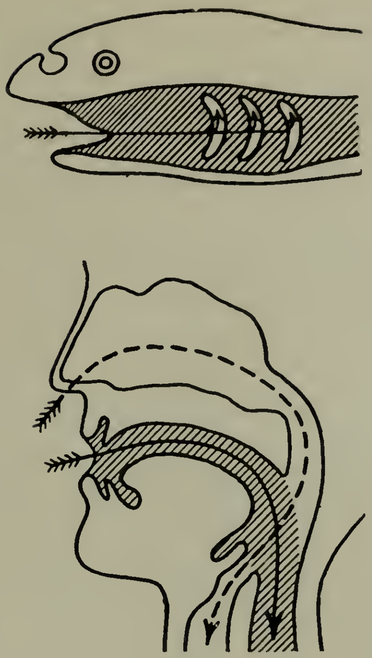


FIGURE 3

MIDSECTION OF THE HEAD OF A FISH AND THE HEAD OF A MAN TO SHOW THE NEW RELATIONS OF THE WATER COURSE FROM THE MOUTH TO THE GILL REGION IN AIR-INHABITING ANIMALS

Note that in the pharynx the air current passes to and fro across the ancient water course (which is cross-hatched).

mouth and the pharynx consists of flattened epithelium in which there are very few of the secreting cells. The air which passes to and fro across this ancient water course would therefore tend specifically to dry the region. Prolonged speaking, singing, or smoking, which would deprive the pharynx of the advantage of previously moistened air, because the air would be drawn in mainly through the mouth and not through the nose, would be especially favorable to drying the mucous membrance at the crossing. The sensation of dryness and stickiness in this area, as already noted, is commonly recognized as thirst.

The question now arises, why does not the mucous membrance of the back of the mouth always feel dry and sticky? And since it does not always feel so, why does it feel so when the body is in need of water? Again, a comparison of water-inhabiting animals with air-inhabiting animals offers a suggestive clue. The study of comparative anatomy has revealed the fact that vertebrates which live in water do not possess buccal glands. Vertebrates which live in air, however, have such glands as newly developed features associated with their new type of environment. In the higher air-inhabiting forms the buccal glands have become evolved into the well-known salivary glands. There are three of these on each side of the mouth, each gland provided with a duct. The ducts empty either on the side of the mouth or under the tongue. The saliva which the glands produce has a water content varying between 97 and 99 per cent. It is a striking fact that mammalian forms which have returned to a water habitat after having developed as land animals, such as the whales and the porpoises, have no salivary glands or very small remnants. The theory which these facts suggest is that when water is lacking in the body the salivary glands are unfavorably affected, along with other structures, by the deficient water supply; that they differ from other structures, such as muscle, for example, in requiring a large amount of water for the performance of their function, which is that of pouring out a secretion consisting almost wholly of water; and, furthermore, that they occupy a peculiarly strategic position, for if they do not have water to utilize for secretion and are therefore unable to secrete, the mouth and pharynx become dry and thus the sensation of thirst arises. Such is the theory of thirst which would account for a local, peripheral source at the back of the mouth, when there is general bodily need of water.

Evidence that Thirst is of Local Origin. The evidence that thirst results from a dryness due to deficiency of saliva may now be given:

1) The chewing of a tasteless gum for five minutes will result repeatedly in a fairly uniform amount of saliva. In observations which Cannon (12) made upon himself this averaged about 14 cc. In an experiment in which no fluid was taken between 7:00 P.M. of one day and 3:00 P.M. of the next day, there was no diminution of the amount of saliva that resulted from the standard period of mastication until after 11:00 A.M. During the next three hours of the test the amount secreted was gradually reduced from approximately 14 cc. to less than 8 cc. Thereupon, at 3:00 P.M., 1000 cc. of water were drunk. A collection of saliva during each of the following four hours by the standard method proved that the discharge from the salivary glands had been promptly restored to approximately the previous amount and was there maintained. During the period when the saliva flow was diminishing, the sensation of thirst became prominent. After the water was drunk and the saliva was again secreted, the sensation of thirst disappeared. The coincident association of bodily need for water, diminished flow of saliva, and the sensation of thirst strongly suggests that the deficient functioning of the salivary glands signals the bodily need by causing the unpleasant sensation.

2) By wrapping the body in very warm blankets and applying hot water bottles it is possible to cause abundant sweating. The large loss of water from the body induced thereby was shown to lessen the output of saliva, which results from standard stimulation for a standard period, to about one-half what it was before the water loss. Associated therewith was a noteworthy dryness of the mouth and an unpleasant thirstiness. This condition was soon obviated by the drinking of water.

3) In another experiment Cannon (12, chap. XVI) showed that the subcutaneous injection of the drug atropine caused the salivary output, resulting from mastication for a standard period, to fall from 13.5 cc. to 1 cc. This occurred without any noteworthy loss of water from the body. Nevertheless, all the feelings of ordinary thirst were present. The unpleasant dry surface of the interior of the mouth, the sense of stickiness,

the difficulty of speaking and swallowing—all these particular features of thirst appeared in association with the characteristic mass sensation. In this experiment atropine had its usual peripheral effect of stopping the flow of saliva; and by thus producing local dryness of the mouth it gave rise to the usual experience which attends that condition.

4) There is a well-established reflex secretion of saliva after the mouth becomes slightly dry. A very simple experiment is that of chewing an indifferent substance for five minutes and comparing the amount of saliva thus obtained with that which is obtained by breathing only through the mouth for five minutes. At first the passage of air to and fro through the mouth gradually dries the surface. As the surface becomes more and more dry, however, saliva is poured out and the amount collected as it flows forth may be considerably greater than that obtained by the chewing. The presence of this reflex indicates that salivary glands have as one of their special functions the moistening of the mouth. This observation is in harmony with the observations made by Pavlov (18) that dry food is peculiarly effective in evoking a marked salivary discharge.

5) Many years ago Bidder and Schmidt (11) tied in dogs the ducts of the salivary glands and made the interesting observation that such animals are always ready to drink water. The tying of the ducts, of course, has no effect on the water content of the body, but it is effective, like atropine, in causing a dryness of the mouth. Under these circumstances the animal acts as if quite as thirsty as an animal long deprived of water. The local dryness fully accounts for the observed behavior.

6) Thirst as a consequence of fright and the attendant checking of salivary secretion are well known. Dr. H. J. Howard (14), of Peking Union Medical College, has vividly reported his experiences when he thought he was about to be shot by Chinese bandits. "So I was going to be shot like a dog! My tongue began to swell, and my mouth to get dry. This thirst rapidly became worse until my tongue clove to the roof of my mouth, and I could scarcely get my breath. The thirst was choking me . . . I was in a terrible state of fear." He prayed for strength to meet his approaching doom and soon fear left him as he determined to die like a man. "Instantly my thirst began to disappear," he writes. "In less than a minute it was entirely gone and by the time we had reached the gate I was perfectly calm and unafraid." Again the intense and distressing sensation of thirst was not associated with a real lack of fluid in the body but resulted from a local condition in the mouth.

7) Pack (17) has reported that he deprived rabbits of water for seven days. Thereupon he gave some of them the drug pilocarpine, which, even when the water supply of the body is low, causes a secretion from the salivary glands. He then allowed all of the animals to drink. Those which were salivating because of the action of pilocarpine either refused to drink or in two cases drank within an hour only 15 and 25 cc. of water. Within a half hour the others drank from 62 to 137 cc. The difference in behavior of these animals can be accounted for by the moistness of the mouth due to salivary secretion in the animals which did not drink and by the dryness of the mouth in the others. In all the animals the general need

of the body was equally great, for they had been without water for seven days.

The foregoing observations taken together support the conclusion that thirst is normally the consequence of a drying of the mucous membrane of the mouth and pharynx when the salivary glands fail to keep this region moist. The continuous loss of water from the body through the kidneys, through the respiratory passages and the skin, does not cause any appreciable change in the water content of the *blood* for a long period. Observations made by Mayer (15) have shown that there may be no demonstrable change in the blood of a dog after three days of total deprivation of water. It is clear that the blood is maintained in a constant state at the expense of the water reserves in the tissues and in the cells of the body structures. Among the structures which are called upon are the salivary glands. As pointed out above, however, they require water for their proper service to the organism. Not having water available they cannot perform that service and consequently the mouth becomes disagreeably dry. When water is drunk it is at once made available for the salivary glands and they can again perform their special function of keeping the buccal cavity moist and lubricated.

The evidence cited above also offers a new explanation of the experiments which have been adduced to support the theory of thirst as a general sensation. It is obvious that if a condition of thirst has arisen because of bodily need for water, the introduction of water under the skin or into a blood-vessel or by way of the intestines will result in a supply which will allow the salivary glands to operate. The mouth, which has been dry because of their failure to operate properly, will be moistened again, and the sensation of thirst will disappear. Its disappearance will not result directly from the satisfaction of a general bodily need but will result indirectly from the moistening of the buccal lining, which, when dry, produces a highly disturbing experience.

In the foregoing discussions of the nature and physiological basis of hunger and thirst no attention has been given to the abnormal manifestations of these two states. There are instances of intense hunger beyond normal limits and also instances of unquenchable thirst as, for example, in diabetes insipidus. Such pathological states are of interest in connection with the physiological explanations for these sensations, but space will not permit discussion of them in this brief account.

The modes of action of the appetites for food and drink and of the sensations of hunger and thirst in maintaining the bodily supplies of nutriment and water may be regarded as typical of other arrangements in the organism which operate for the welfare of the individual or the race. As Perry (19) has pointed out, behavior may be directed by both negative expectation—the riddance of disturbing stimulation—and by positive expectation—the prolongation or renewal of agreeable stimulation. Hunger and thirst belong to the first category. Also in this category may be classified the asphyxia of air-hunger, the tensions of "sex hunger," pain, and possibly also the discomfort of fatigue and the unpleasantness of confinement or physical restraint. Each of these states is associated with an impulsive factor; each one more or less vigorously spurs or drives to action; each may be so disturbing as to force the person who is afflicted to seek relief from the intolerable annoyance or distress. On the other hand, experience may condition behavior by revealing that the taking of food, drink, or exercise is accompanied by unanticipated delight. Appetites for the repetition of these experiences are thus established; the person beset by an appetite is tempted, not driven, to action—he seeks satisfaction, not relief. It is not to be supposed that the two motivating agencies—the pang and the pleasure—are as separate as they have been regarded in the foregoing discussion. They may be closely mingled; when relief is found, the appetite may simultaneously be satiated. In so far as an assurance of supplies for food and water is concerned, appetite, or the habitual taking of these provisions, is the prime effective agency. If the requirements of the body are not met, however, in this mild and incidental manner, hunger and thirst arise as powerful, persistent, and tormenting stimuli, which imperiously demand the taking of food before they will cease their goading.

REFERENCES

Section on Hunger

1. Bulatao, E., & Carlson, A. J. Influence of experimental changes in blood sugar level on gastric hunger contractions. *Amer. J. Physiol.*, 1924, **69**, 107-115.
2. Cannon, W. B. The mechanical factors of digestion. London: Edward Arnold, 1911. Pp. 227.
3. ————. Bodily changes in pain, hunger, fear, and rage. New York: Appleton, 1929. Pp. 404.
4. Cannon, W. B., & Washburn, A. L. An explanation of hunger. *Amer. J. Physiol.*, 1912, **29**, 441-454.
5. Carlson, A. J. A study of the mechanisms of the hunger contractions of the empty stomach by experiments on dogs. *Amer. J. Physiol.*, 1913, **32**, 369-388.
6. ————. The control of hunger in health and disease. Chicago: Univ. Chicago Press, 1916. Pp. 319.
7. ————. The relation between the contractions of the empty stomach and the sensation of hunger. *Amer. J. Physiol.*, 1912, **31**, 175-192.
8. Hoelzel, F. Central factors in hunger. *Amer. J. Physiol.*, 1927, **82**, 665-671.
9. Müller, L. R. Die Lebensnerven. Berlin: Springer, 1924. Pp. 614.
10. Rogers, F. T., & Martin, C. L. X-ray observations of hunger contractions in man. *Amer. J. Physiol.*, 1926, **76**, 349-353.

Section on Thirst

11. Bidder, F., & Schmidt, C. Die Verdauungssäfte und der Stoffwechsel. Leipzig: Reyher, 1852. Pp. 413.
12. Cannon, W. B. Bodily changes in pain, hunger, fear and rage. New York: Appleton, 1929. Pp. 404.
13. ————. The physiological basis of thirst. *Proc. Roy. Soc.* (*B*), 1918, **90**, 283-301.
14. Howard, H. J. Ten weeks with Chinese bandits. New York: Dodd, Mead, 1926. Pp. 272.

15. Mayer, A. Essai sur la soif. Paris: Jouve Boyer, 1900. Pp. 168.
16. Müller, L. R. Die Lebensnerven. Berlin: Springer, 1924. Pp. 614.
17. Pack, G. T. New experiments on the nature of the sensation of thirst. *Amer. J. Physiol.,* 1923, **65**, 346-349.
18. Pavlov, I. P. The work of the digestive glands. London: Griffin, 1910. Pp. 266.
19. Perry, R. B. General theory of value. New York: Longmans, Green, 1926. Pp. 702.

CHAPTER 12

EMOTION: I. THE NEURO-HUMORAL BASIS OF EMOTIONAL REACTIONS

PHILIP BARD
Princeton University

For centuries bodily change has been the criterion of emotion. The rôle of the musculature of limbs, trunk, and face in producing the attitudes, expressions, and movements typical of the several emotions has been evident at all times. At a very early date observers began to pay special attention to the more obvious visceral activities which accompany affective states. Where little was known of the brain and its relation to the internal organs there was a natural tendency to locate the feeling of emotion in the reactive viscera. In the Book of Genesis we read that when Joseph recognized Benjamin "his bowels did yearn upon his brother." During Homeric times and throughout early Greek medicine psychic functions were generally attributed to thoracic and abdominal organs. Aristotle, who refused to attach much importance to the brain, made significant observations on the heart and by a reasonable process of thought came to regard that organ as the principal seat of psychic activity. To the influence of Aristotle and the writers of the Old Testament we largely owe such expressions of common parlance as "matters of the heart," "hard-hearted," and "bowels of mercy."

Galen (131-201 A.D.), whose authority dominated anatomical and physiological thought for fifteen centuries, first showed experimentally that voluntary movement and sensation are dependent upon the connection of the peripheral nerves with the brain. When Vesalius (1514-1562), the great founder of modern anatomy, confirmed and amplified many of Galen's neurological findings, it was definitely recognized that the brain is the seat of sensation and the organ to which voluntary movement must be referred. In addition, the brain, as the sensorium for the whole body, became, at the hands of Descartes, the site of emotional consciousness. It was then only a step to assume that emotional behavior is dependent upon cerebral processes.

It was easy enough for the writers who followed Vesalius to postulate a cerebral origin for the activities of skeletal muscles in emotion. To them these changes were obviously due to influences transmitted by the same cerebrospinal nerves which execute the dictates of the will, for they may be modified by the will acting through the same channels or may be mimicked by voluntary effort in the absence of any affective state. But in the case of the wholly involuntary visceral accompaniments of emotion the neural basis seemed less clear. To be sure Galen had described visceral nerves (53) and Eustachius, the contemporary and rival of Vesalius, portrayed them in a way that almost rivals modern presentations of the subject, but at that time their relation to the brain was unknown or incorrectly surmised. Yet the necessity for some connection between brain and viscera was impressed upon all who gave thought to the visceral reverberations of emotional excitement. In the light of recent work it is interesting to see how often during the seventeenth and eighteenth centuries the opinion was expressed that the influence which the emotions exert upon the internal organs is mediated by that peripheral group of nerves which is now called the sympathetic system. Certain writers of that period went so far as to locate the central control of the viscera in definite parts of the brain. Thomas Willis (1621-1675), who wrote the most complete account of the central nervous system of his time, placed it in the cerebellum. But these early conjectures, historically interesting as they are, lacked experimental support and it was not until the precise central origins, the general distribution and the particular peripheral effects of the visceral nerves were worked out during the second half of the last century that a neural basis for the visceral aspects of emotional behavior could be accurately described.

The somatic[1] manifestations of such elementary states as fear and rage include external movements of offense and defense which are effective in preserving the organism at times of stress. Their utility has always been evident. But insight into the significance of the visceral alterations which occur concomitantly with the overt action has come only with recent physiological progress. On the basis of ample experimental findings we now recognize in them a cooperation of the viscera which is directly serviceable to the organism in preparing for and maintaining the appropriate external activities. The peripheral nervous mechanism for the somatic expression of emotion is the same as that involved in voluntary and reflex activities of the skeletal muscles. It consists of the familiar somatic components of the cranial and spinal nerves. On the other hand the viscera are connected with the central nervous system in a special manner, and it will be necessary to devote some attention to the general organization and functional importance of the visceral innervation before considering its rôle in producing emotional behavior. It will then be possible to turn to the central nervous mechanisms which activate both sets of peripheral paths during emotional excitement.

The Visceral Nerves Concerned in Emotional Behavior

The motor neurons which connect the central nervous system with skeletal muscle-fibers have their cell bodies in the brain or spinal cord; one nerve-cell bridges the gap between cerebrospinal axis and effector. The innervation of the secreting cells and the smooth muscle of the viscera is somewhat different. It is characterized by the fact that the axones which come into immediate functional relation with these effector cells belong to neurons which lie entirely outside the central nervous system. The cell bodies of these outlying neurons are typically grouped together in peripheral ganglia. There they make synaptic connections with axones arising in the central nervous system. The special feature of the innervation of the viscera lies in this fact that the path from central axis to effector is broken by a synapse. This arrangement is illustrated in Figure 1, which shows the essential difference between the somatic and visceral innervations. The neuron on the central side of the peripheral synapse is called *preganglionic* and that on the distal side *postganglionic.* From embryological studies we know that the postganglionic neurons originate in the central nervous system and, in the course of development, migrate outward to take their characteristic peripheral positions. They are thus the homologues of the motor fibers of the cerebrospinal system, while the preganglionic neurons have their counterparts in the centrally placed premotor neurones of that system. In Figure 1 this homology is illustrated by the use of broken lines.

---

[1] In the absence of a better term the word "somatic" is used throughout this chapter to describe the activities and innervation of skeletal muscle as contrasted with those of the viscera.

One other morphological feature distinguishes the visceral from the somatic innervation. While the cerebrospinal fibers pass out from the central nervous system in an almost unbroken series, the preganglionic fibers issue in three distinct groups, as was first shown by Gaskell (32). The first of these, the *cranial* outflow, is composed of visceral motor fibers in the third, seventh, ninth, and tenth pairs of cranial nerves. The fibers of the second division arise from all the thoracic and from the upper lumbar segments of the spinal cord and begin their peripheral course in the ventral roots of the corresponding spinal nerves; they form the *thoraco-lumbar* outflow. The third division, the *sacral* outflow, consists of fibers which proceed from the sacral spinal segments in a manner similar to the mode of origin of the thoraco-lumbar division. Visceral fibers do not issue in some of the cranial nerves, and the cervical and lower lumbar segments of the cord do not contribute to the preganglionic outflow (see Figure 2).

The entire visceral nervous system, postganglionic neurons and ganglia as well as preganglionic neurons, may be divided along the lines of these three divisions. Unfortunately the names applied to the three parts have varied from author to author and few are appropriate. The entire system composed of preganglionic and postganglionic neurons was called by Langley, to whom we owe much of our knowledge of these structures, the *autonomic nervous system* (44). He used for the thoraco-lumbar division the old name *sympathetic* and classed the cranial and sacral divisions together under the term *parasympathetic.* This terminology is the one commonly used in current physiological literature and will be followed in the present discussion. The word sympathetic has been applied to each and all of the three divisions, but was originally introduced in 1732 by Winslow to describe the chain of ganglia and associated strands which we now know to be connected with the thoracic and lumbar segments of the spinal cord. Accordingly, Langley's use of the term has a historical justification. Since the cranial and sacral divisions possess certain common features which distinguish them from the middle division, they are conveniently described together as the parasympathetic system.

Strictly speaking, the autonomic system is made up entirely of motor nerves. Langley (43) was unable to obtain any experimental evidence of a true reflex being mediated through autonomic ganglia. There is no good evidence that these structures ever act as reflex centers for the nervous reactions of the viscera. Although visceral afferent fibers are found in many of the nerves of this system they have their cell bodies in the sensory ganglia of the cranial and spinal nerves and so belong to the cerebrospinal system rather than to the autonomic.

The Sympathetic Division of the Autonomic System. The preganglionic fibers of this division have their cell bodies in the lateral horns of the gray matter of the thoracic and upper lumbar spinal segments (32).

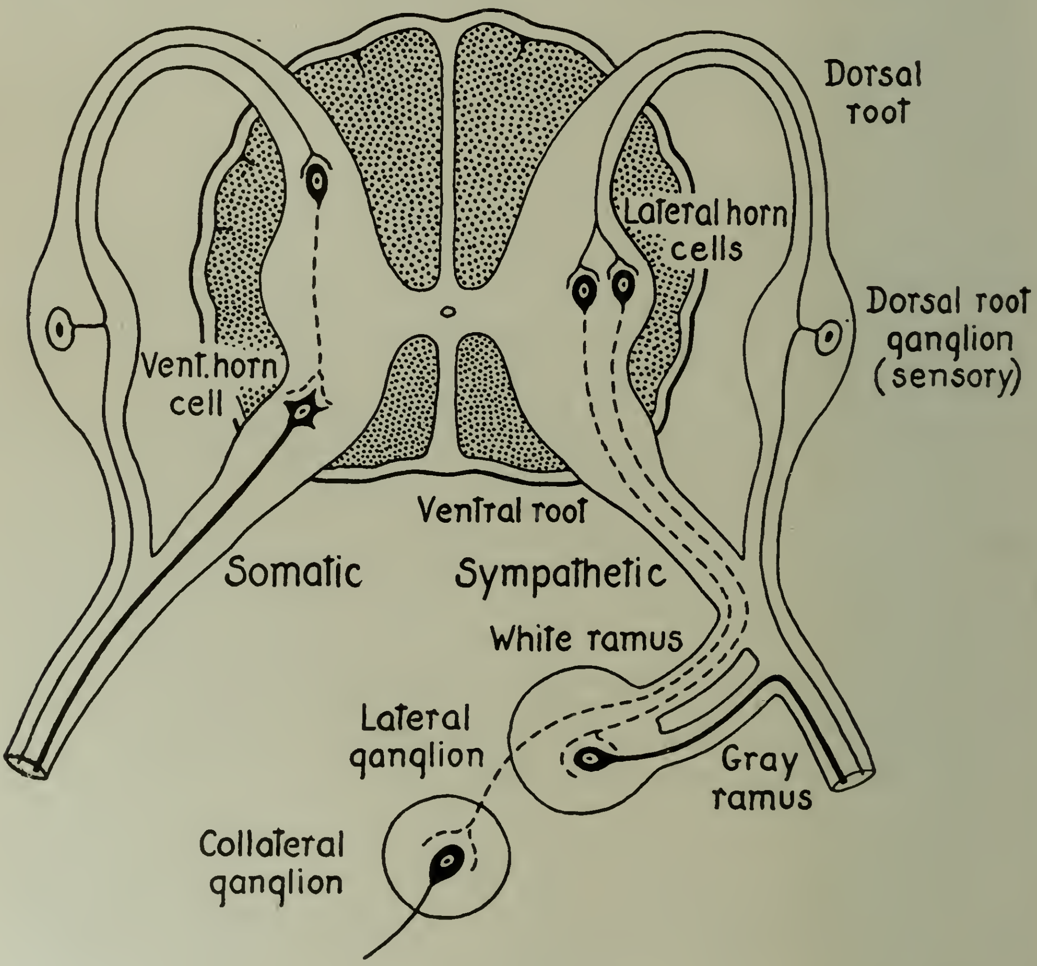


FIGURE 1

Diagram Illustrating the Different Arrangements of Neurons in Somatic and Autonomic Nervous Systems

At left is shown simple spinal reflex arc of somatic system, at right that of sympathetic division of the autonomic system. At right the preganglionic neurons are represented by broken lines, the postganglionic by solid lines. (After Gaskell. From: MacLeod, J. J. R. Physiology and biochemistry in modern medicine. [5th ed.] St. Louis, Mo.: Mosby, 1926. Fig. 56.)

Emerging with the somatic motor fibers in the ventral roots, they leave the spinal nerves to form the white rami communicantes and course to the ganglia of the sympathetic system (see Figure 1). These ganglia fall into two distinct groups.

The lateral (or vertebral) ganglia are arranged in two chains, one on either side of the spinal column. These chains extend from the upper part of the neck to the pelvic region. In general, their ganglia have a segmental distribution, but in the upper thoracic and in the cervical regions several have united to form the large stellate and superior cervical ganglia (Figure 2). One characteristic of the vertebral ganglia is that they send postganglionic fibers to the spinal nerves. These fibers form the gray rami communicantes. Gray rami contribute to the composition of every spinal nerve while the white rami arise only from the thoracic and upper lumbar ventral roots. By means of the gray rami and spinal nerves, postganglionic fibers are carried to the blood-vessels of skin and skeletal muscle, to the sweat-glands, and to the smooth muscle of the hairs. Postganglionic fibers from the superior cervical ganglia join the fifth and one or two other pairs of cranial nerves to be distributed to the smooth muscle of the orbit and to the glands and blood-vessels of the cranial region. From some of the lateral ganglia, notably those of the upper thorax, postganglionic fibers run directly to the organ innervated. This is so in the case of the cardiac accelerator nerves (18).

The collateral (or prevertebral) ganglia are the ganglia of the so-called solar plexus (the semilunar and superior mesenteric ganglia), the inferior mesenteric ganglia and some small nearby ganglia. These lie near the points of origin of the main arteries supplying the abdominal organs. From them postganglionic fibers proceed with the arteries to innervate the abdominal viscera and their blood-vessels. These ganglia receive preganglionic fibers through white rami from the fifth thoracic to the second lumbar nerves. In this case the preganglionic fibers join the lateral chain, but continue to the collateral ganglia as the *splanchnic* nerves.

The fibers of the sympathetic system have a very wide distribution throughout the body. They include vasoconstrictor fibers to the small blood-vessels of the abdominal organs, skin, and other parts, accelerator fibers to the heart, secretory fibers to the sweat-glands, inhibitory fibers to the smooth muscle of the gastro-intestinal tract, motor fibers to smooth muscle of the hairs, fibers supplying the liver which carry impulses that act to release sugar into the blood, constrictor fibers to the spleen, dilator fibers to the bronchioles, cervical sympathetic fibers to the eyes causing dilatation of the pupils and retraction of the nictitating membranes (where these are developed) and projection of the eye forward in the orbit (exophthalmos), fibers to the pelvic viscera, and finally, secretory fibers to the medullary portion of the adrenal glands.

The innervation of the adrenal medulla is peculiar. Elliott (29) has shown that the secretory cells of this endocrine organ receive preganglionic fibers directly from the splanchnic nerves. This apparent exception to the rule that organs innervated by the autonomic system receive postganglionic fibers is explained by the fact that the secreting cell of the adrenal medulla

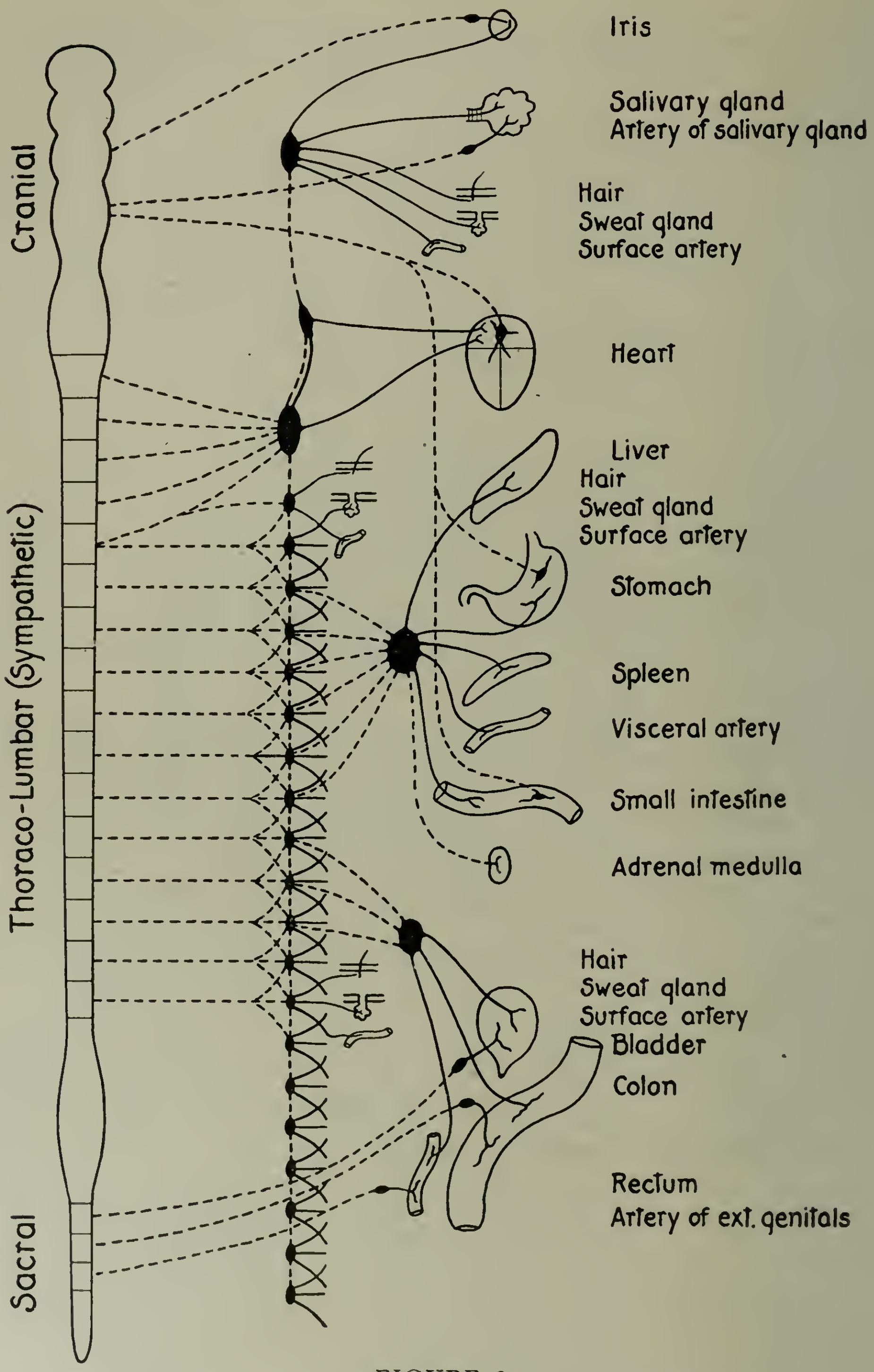


FIGURE 2

Diagram of the General Arrangement of the Autonomic Nervous System

The brain and spinal cord are represented at left. The nerves of the somatic system are not shown. The preganglionic fibers are in broken lines, the postganglionic in solid lines. Further description in text. (After: Cannon, W. B. Bodily changes in pain, hunger, fear and rage. New York: D. Appleton and Company, 1915. P. 24.)

is the morphological equivalent of the postganglionic sympathetic neuron. In the embryo both types of cells arise from a common central mass and in certain lower vertebrates they have a common peripheral distribution (32). It is not surprising, therefore, that the internal secretion of these glandular cells, adrenin, when circulating in the blood stream, acts upon any tissue in the same way as do nerve-impulses conducted by sympathetic postganglionic fibers (44).[2] These facts make it clear that the adrenal medulla is an integral part of the sympathetic system.

Cannon (11) has pointed out that besides the *extensiveness* of the distribution of its fibers another important feature of the sympathetic system is its arrangement for a *diffuse* discharge. It was found by Langley (43) that the sympathetic preganglionic fibers issuing from any one spinal segment pass to a series of lateral ganglia, usually from six to nine in number, never to a single segmental ganglion, and he also obtained evidence that each preganglionic fiber sends branches to make synaptic connections with postganglionic neurons in several ganglia. This is a morphological condition which assures a widespread discharge over postganglionic fibers. Furthermore, a discharge of impulses over that single group of preganglionic fibers which innervates the adrenal medulla will throw into the blood stream a chemical agent capable of inducing practically all of the changes which occur when the entire thoraco-lumbar outflow is active. The existence of these neuro-humoral arrangements for the production of extensive effects leads plainly to the conclusion that the sympathetic division of the autonomic system is organized to *act as a whole,* a view that has been particularly emphasized by Cannon (11).

This system of nerves has many duties in the animal economy. A certain fraction of the preganglionic outflow is constantly discharging and this tonic activity is in part responsible for a number of fairly constant bodily conditions. A continuous activity of vasoconstrictor fibers keeps the smooth muscle of the arterioles in a state of sustained contraction, a peripheral condition of major importance in maintaining an adequate arterial pressure. The pulse rate tends to remain relatively rapid because of the tonic discharge of cardio-accelerator fibers. Section of the cervical sympathetic trunk leads at once to constriction of the pupil, sinking of the eye into the orbit (enophthalmos), and, in animals which possess an active nictitating membrane, passive projection of that tissue over the front of the eye. This result demonstrates that a tonic discharge over cervical sympathetic fibers is responsible for the normal size of the pupil and for the normal position of the eye in the orbit. These tonic sympathetic discharges are of central origin. They may be decreased by inhibition or augmented by excitation of the central neural mechanisms in control of the preganglionic neurones. Certain sympathetic paths such as those to the sweat-glands and smooth muscle of the hairs appear to be usually at rest and become active only when special conditions lead to excitation of their central connections. It is also true that one or more functional groups of sympathetic fibers may be affected without the involvement of other groups. A lowered arterial pressure may lead to compensatory augmented activity of the vasoconstrictor and cardio-accelerator mechanisms without extension of the response to other parts of the sympathetic system. Several instances of similar fractional activity in the sympathetic might be given. But the most general and the most significant aspect of the physiology of this system is its marked tendency to go into action as a whole.

---

[2] The sweat-glands have a sympathetic innervation, but are not affected by adrenin (44). This is the only known exception to the statement made above.

Simultaneous activity of all parts of the somatic system of nerves would lead to a generalized muscular spasm such as occurs in strychnine poisoning. Under the influence of this drug the normal integrative action of the nervous sytem gives way to a diffuse discharge of impulses and mass incoordinate movements result. Does action as a whole in the sympathetic evoke similar disintegrated activity or do the many changes so induced combine to form a purposive effect? Physiological consideration of these changes shows the majority of them to be related in such a way that they cooperate to mobilize energy and to provide for its ready utilization. Energy transformation in the body involves an oxidation of foodstuffs with the production of carbon dioxide, other metabolites, and heat. Its principal seat is the skeletal musculature. When it is augmented, the active tissues demand more food, more oxygen, and the means of distributing and disposing of additional quantities of metabolites and heat. These requirements must be met by prompt internal responses. Activity in the sympathetic system is the immediate cause of many of them. It leads to an increase of the available potential energy by throwing more sugar into the blood. It stops unnecessary activity in the digestive system. By accelerating the heart and causing vasoconstriction in quiescent regions it produces a more rapid circulation through the active tissue. It thus provides a means of supplying additional food and extra oxygen to the site of augmented metabolism and makes possible an adequate removal of metabolites and heat. Sympathetic activity further facilitates the respiratory processes by dilating the bronchioles, and recent work (37) has given evidence that it may discharge into the blood an extra supply of the oxygen-carrying red corpuscles by causing contracton of the spleen. Through their action on the sweat glands sympathetic impulses provide for loss of excessive amounts of heat. Secreted adrenin not only supplements the effects of sympathetic impulses but is capable of acting as an antidote to the effects of fatigue in skeletal muscle (11). It is therefore clear that when circumstances demand an increased energy output a tendency on the part of the sympathetic to discharge as a whole will lead to bodily changes which constitute a purposive reaction.

Cannon (11, 14) has pointed out that the conditions under which the sympathetic actually does discharge as a unit are those which require a vigorous response to insure the maintenance of an essential bodily state or even of life itself. Such conditions include strong emotional excitement, pain, vigorous muscular exercise, asphyxia, exposure to cold, and states of lowered blood sugar (hypoglycemia). From the account just given it can be seen that in muscular activity with its increased energy output a diffuse and widespread discharge of sympathetic impulses would go far in mobilizing bodily resources. Since pain, asphyxiation, and strong emotion are likely to be accompanied or followed by muscular effort and struggle, the visceral changes evoked by sympathetic activity are appropriate in those states. When the warm-blooded animal is exposed to a lowered environmental temperature, reactions occur which prevent the threatened depression of body temperature. Restlessness and shivering produce additional quantities of heat. Activity in the sympathetic by giving extra blood sugar provides a readily combustible material for the extra heat production; by causing constriction of peripheral vessels and erection of hair it diminishes the heat loss from the body surface and by augmenting secretion of adrenin it supplies a substance which not only reinforces these changes but is capable of hastening combustion itself (21). The blood sugar, like the blood temperature, is kept at a fairly constant level. Normally it lies around 100 mg. per 100 cc. of blood. When this percentage is greatly diminished as by an excessive dose of insulin, so-called hypoglycemic reactions occur. They first appear when the sugar falls below 70 to 80 mg. and include pallor, a rapid pulse, dilatation of the pupils, and profuse sweating. Cannon, McIver, and Bliss (20) have shown that with these signs of sympathetic activity there is an increment in medulliadrenal secretion of sufficient magnitude to check the rate of fall of the glycemic percentage. Since convulsions occur, and coma and death may ensue when the blood sugar reaches 45 mg., this sympathico-adrenal activity is of prime importance in warding off a dangerous deficiency. In short, each of these six situations constitutes a state of stress, an emergency, in which dire consequences threaten the integrity of the organism. In each, widespread sympathetic activity leads to changes which anticipate or ameliorate the danger. We owe the realization of this important fact to Cannon who has described it as the *emergency function of the sympathico-adrenal system* (11, 12, 14).

This distinctive function of the mechanism naturally raises the question of its indispensability. It may be said at once that there is abundant evidence that the adrenal medulla, unlike the adrenal cortex, is not essential for life or even normal existence (50). Recently Cannon, Lewis, and Britton (19) have successfully removed from cats both lateral sympathetic chains from superior cervical to coccygeal ganglia. After degeneration of the truncated postganglionic fibers animals thus operated upon possess no functional sympathetic nervous tissue. A number of such animals have lived in good health for over a year and so long as they are kept under quiet laboratory conditions they show no deficiencies of behavior. But when these "pansympathectomized" cats are confronted with a stressful situation, they are clearly at a disadvantage in comparison with their unoperated fellows. In cold weather, they lose heat rapidly, have a low body temperature, and they seek a warm place more than do normal cats. When excited their blood sugar does not rise and the hair is not erected. They lack any control over secretion of adrenin. Although they exhibit all the typical somatic signs of aggressive feeling when confronted with their enemy, the dog, they cannot summon the visceral support which is under sympathetic control. They would enter combat or take flight under serious disadvantages. These findings at once demonstrate the dispensability of the sympathetic and emphasize its emergency function.

The Parasympathetic Division of the Autonomic System. This second part of the autonomic, formed by classifying together the cranial and sacral divisions, is composed of fibers which bear nerve-impulses to many organs situated in diverse parts of the body. The third cranial nerves supply motor impulses to the ciliary muscle and the sphincter of the pupil. The seventh and ninth nerves distribute secretory and vasodilator impulses to the salivary glands and the mucous membranes of the buccal region. The tenth pair of cranial nerves, the vagi, furnish inhibitory fibers to the heart and motor fibers to the glands and musculature of the gastro-intestinal tract. Turning to the sacral outflow we find the pelvic nerves (nervi erigentes) composed of fibers which cause contraction of colon, rectum, and bladder, and vasodilation in the external generative organs.

In the variety of its distribution and in the number of effects which it evokes, the parasympathetic resembles the sympathetic. Indeed, a majority of the viscera are innervated by both systems. But here the likeness ends. The most striking fact about the organization of the parasympathetic is that groups of preganglionic fibers reach out to ganglia which lie close to or within the organ innervated. The postganglionic fibers issuing from any one ganglion are short and supply a circumscribed mass of tissue (Figure 2). This *restricted distribution* is an arrangement whereby the central nervous system is able to exert effects upon single organs without at the same time affecting other organs. It is this feature which distinguishes the parasympathetic most sharply from the sympathetic where the organization is favorable to the production of simultaneous activity in a large number of viscera.

In accord with this fact is the absence of any parasympathetic hormone capable of acting extensively upon organs innervated by this system. The adrenal medulla has no counterpart in the parasympathetic. The reason for this is clear. Action as a whole in this instance would lead to widespread coincident changes having no functional relation to one another; the result would be disorderly and purposeless. Constriction of the pupil, slowing of the heart, increased gastro-intestinal activity, contraction of the bladder and rectum, and engorgement of erectile tissue are not responses which are bound by any physiologic ties of integration.

Despite the lack of functional interrelation between the several groups of parasympathetic fibers, it is possible to describe in a general way the respective functions of the cranial and sacral divisions. Cannon (11) has pointed out that the former may be regarded as a conserver of bodily resources. The tonic action of the cardio-inhibitory nerves under normal conditions and their augmented discharge when the heart is overtaxed by a high arterial pressure protect that organ from overactivity. Similarly, constriction of the pupil by reflex activation of third nerve-fibers protects the retina from excessive light and even under ordinary intensities of illumination a tonic discharge of these parasympathetic fibers is a necessary condition for acute vision as anyone who has experienced the paralytic effects of atropine ("drops in the eye") will testify. The conserving function of the cranial autonomic is perhaps best illustrated by its control over digestive processes. The secretory fibers of the seventh and ninth nerves are responsible for the secretion of saliva; vagal fibers, for much of the normal activity of oesophagus, stomach, and small intestines. In this way the cranial parasympathetic contributes in a very definite manner to the nutrition of the body.

The sacral autonomic has been characterized (11) as a group of mechanisms for emptying. Normally the contractions of bladder, colon, and rectum are reflex occurrences in which the adequate stimulus is distention by accumulation of contents. But it is notorious that these organs may be emptied during those strong emotional states which activate the sympathetic division, and we must recognize that vigorous sympathetic discharge may be accompanied by discharge via sacral autonomic fibers. The rhythmic contractions which empty the vasa deferentia and seminal vesicles at the height of sexual excitement and the supposedly analogous contractions of the uterus are not due to a sacral discharge—these organs are supplied only by fibers of sympathetic origin (43). However, a discharge of nerve-impulses over sacral vasodilator fibers to the external genitalia of both sexes is the usual forerunner of these events and in the male it is commonly evoked by distention of the tubes and vesicles.

The Antagonism of the Two Divisions of the Autonomic System. Certain structures, namely, the internal generative organs, the blood-vessels and glands and smooth muscle of the skin, the blood-vessels of the portion of the digestive tract between mouth and rectum and of the glands opening into that portion, receive autonomic fibers only from the sympathetic system. Most of the visceral tissues, however, possess a double nerve supply, sympathetic and cranial or sympathetic and sacral (43). Wherever this occurs, the sympathetic and parasympathetic as a rule produce opposed effects. In the heart the sympathetic exerts a positive, the cranial a negative influence upon excitability, conduction and rate and strength of beat (56). In general, it can be said that the sympathetic sends inhibitory nerves to the smooth muscle of the entire gastro-intestinal tract except the sphincters. The parasympathetic, on the other hand, supplies motor fibers to the stomach and small intestine by way of the vagi and to the colon and rectum through the sacral outflow. The cranial innervation of the eye contracts the pupil, the sympathetic dilates it. While the parasympathetic sends vasodilator fibers to the salivary glands and external genitalia, these parts, in common with the rest of the body, receive a vasoconstrictor supply from the sympathetic.

The common opposition of the effects of sympathetic and parasympathetic impulses is the basis of certain reciprocal relationships between the functions of the two systems. It is well known, for example, that in the reflex widening of the pupil which follows painful stimulation central inhibition of the constant third nerve discharge to the sphincter of the pupil cooperates with augmented sympathetic discharge to the antagonistic radial muscle of the iris. This is quite similar to the reciprocal innervation of antagonistic skeletal muscles demonstrated by Sherrington (52). In both cases the organization of the central neural mechanisms is such that an increased discharge of impulses to one muscle is accompanied by a diminished discharge to its antagonist. Now many visceral organs, unlike skeletal muscle, receive inhibitory as well as excitatory nerves. Where this occurs, as in the heart and blood-vessels, the two sets of fibers may act in a reciprocal manner upon a single effector. The acceleration of the heart during muscular exercise is the outcome partly of lessened vagal tone and partly of greater accelerator discharge (2).

Autonomic Control of the Viscera in Emotion

Only a little over a century ago Bichat declared (53) that the brain *"n'est jamais affecté dans les passions; les organes de la vie interne en sont le siège unique."* This was a sporadic attempt to revive the ancient doctrine of the peripheral seat of the emotions. It grew out of a belief that the nerves of the viscera lie beyond the control of the central system, that they possess a complete autonomy of function. Such conjectures became untenable as soon as experimental work showed that the normal activity of visceral nerves is dependent upon their connections with the central nervous system and that these same nerves bring about the very changes which constitute the visceral aspects of emotional behavior.

The discovery of vasomotor nerves by Claude Bernard and of the functions of the extrinsic nerves of the heart by German physiologists indicated clearly that the pallor of fear and anger, the blush of shame, and the palpitations of the heart which accompany many emotional states are all due to autonomic fibers. When the secretory nerves of sweat-glands were found to belong to the sympathetic system the "cold sweat" of fear and anxiety had a neural explanation. Erection of hair must have been long noticed as a symptom of strong emotion. It is discussed at some length by Darwin who regarded erection of dermal appendages as perhaps the most general expressive movement common to reptiles, birds, and mammals. In 1870 Schiff (51, pp. 141-146) reported that the phenomenon as seen in the angry cat is abolished by section of sympathetic nerves; later Langley showed conclusively that all pilo-motor nerves belong to the sympathetic system.

Emotion and Digestion. Studies of both the chemical and mechanical factors of digestion have contributed generously to our understanding of the bodily resonance of emotion. They have shown that while certain feelings and emotions are conducive to the secretion of digestive juices and tend to initiate and sustain gastro-intestinal motility, other affective states have a markedly depressive influence upon these processes.

It is a matter of common knowledge that the sight or smell of a favorite food may induce an abundant flow of saliva, a fact described by saying that "it makes the mouth water." This effect is not due to an inborn reflex, but to one that has been acquired in the course of individual experience because of the agreeable quality of the particular food in question (47).

For this reason it is termed a "psychic" or "conditioned" secretion to distinguish it from the unconditioned, unlearned, similar response evoked by the mere presence of food or other material in the mouth. In both types of response the same secretory fibers of the cranial autonomic are activated, in the latter by the reflex machinery of the medulla oblongata, in the former by a more complicated central mechanism involving the cerebral cortex.

That there is also a psychic secretion of gastric juice has been shown by a number of investigators. Pavlov (46), using dogs, has studied the secretion of a side pouch of the stomach prepared surgically in such a way that its cavity was separate from that of the rest of the stomach and opened to the surface of the body. When this arrangement was combined with an opening in the oesophagus through which swallowed food was returned to the outside without reaching the stomach, in short when "sham feeding" could be done, the experimental conditions permitted a study of the effects of presentation, chewing, and swallowing of food upon the flow of gastric juice. It was then found that the sight, smell, or taking of food agreeable to the dogs resulted, after a short latent period, in a copious secretion from the side pouch. Since this result could not be obtained by the mere presence of solid matter or various chemical substances in the mouth and also failed to occur when foodstuffs indifferent to the animals were chewed and swallowed, it was concluded that the adequate stimulus for the response is the pleasure of seeing, smelling, or taking food which is relished. It was further shown by Pavlov that this agency causes stimulation of the gastric glands by discharging impulses over vagal fibers, for after cutting the nerves the swallowing of any amount of highly favored food does not induce the reaction.

While the pleasurable feelings which accompany the anticipation or taking of food promote the initial flow of digestive juices, disagreeable feelings and unpleasant strong emotions have the opposite effect. This has been noted in many studies of gastric secretion and is perhaps best illustrated by the experiments of Bickel and Sasaki (6) who made use of a dog prepared after the manner of Pavlov. Under ordinary quiet conditions a five-minute bout of sham feeding produced, during the following 20 minutes, 66.7 cc. of normal gastric juice. At another time the dog was given a fictitious meal of the same duration after first being infuriated by the presence of a cat. Although he showed every sign of an eager appetite, the procedure yielded only 9 cc. of a thick mucous secretion during the 20-minute period corresponding to the previous observation. It was also found that an abundant psychic secretion which had begun as a result of sham feeding under tranquil conditions was almost wholly stopped by introducing the cat. Quite similar results of similar affective states have been obtained in human cases (1, 11) where injury has led to closure of the oesophagus and made necessary a fistulous opening into the stomach through the side of the body by which food may be introduced.

It is evident from this work that a physiological process accompanying a pleasurable psychic state is inhibited when an unpleasant affective state intervenes. Recently Dumas (26) has reported experiments showing the depressive influence of fear and pain upon gastric secretion. It is curious yet interesting that he should find that sexual excitement, which may be regarded as a strong but not unpleasant emotion, evokes, in dogs, an abundant secretion of saliva and gastric juice.

The motor activities of stomach and intestines are of two types. Rhythmic kneading movements mix the food with the digestive juices and bring the products of digestion in contact with the walls of the gut for absorption. Peristaltic contractions carry the food onward into fresh regions of digestion and absorption and move indigestible residues and excreted matter to a position where they may be voided. In his extensive studies of these mechanical factors of digestion by means of the Roentgen rays, Cannon (10) encountered many instances of emotional influence. Cats which were restive and excited on being fastened to a holder failed to show gastric or intestinal peristalsis, whereas these movements proceeded normally in animals which submitted calmly to the restraint. Whenever signs of rage or distress or mere anxiety made their appearance, there was an immediate cessation of gastro-intestinal movement. The reports of clinical observers are replete with evidence that this depressive effect of emotion lies at the root of many cases of acute and chronic "dyspepsia" (1, 11).

From what is known of the functions of the extrinsic nerves of the gastro-intestinal tract, it must be supposed that the neural basis for the disturbing effect of emotional excitement upon the mechanical aspects of digestion involves a sympathetic discharge. Stimulation of the splanchnic nerves causes diminished tonus and stoppage of contractions in stomach and intestines, and such is the effect of adrenin. On the other hand, the vagi exert a general motor influence. Certain results obtained by Cannon in his work on cats can leave little doubt that the inhibitory effect of emotion upon gastric motility is exerted over splanchnic fibers. It was found that slight interference with breathing, sufficient to evoke an emotional response, caused a complete stoppage of the stomach contractions, provided the splanchnic nerves were intact. After these had been severed and the vagi alone remained, the respiratory distress had no effect upon the stomach. Furthermore, evidence which will be presented in the next section makes it certain that a fit of anger is accompanied by a medulliadrenal secretion capable of producing marked physiological changes, among them depression of the gut.

Some of Cannon's findings have indicated that just as there is a psychic secretion of gastric juice there is also a "psychic tone" of the gastric musculature and that this is mediated by the vagus nerves. The importance of tonic contraction of the stomach muscle lies in the fact that it is necessary for gastric peristalsis. Although it may develop in a stomach disconnected from its extrinsic nerves, vagus stimulation does produce it. These facts raised the question as to the stage in the digestive process at which vagus influences affect gastric tone. It was then found that if the vagal supply to the stomach was cut just before an animal took food, the tonus of the stomach disappeared or remained minimal, and contractions of the gastric wall did not occur. In contrast to this was the effect of cutting the nerves after food had been eaten with relish; then the contractions which had started continued in a normal fashion. The indication here is that the pleasurable taking of food is accompanied by a vagal discharge which arouses the essential tonic condition in the gastric muscles. This may then be continued by local agencies. A recent paper by Alvarez (1) gives abundant testimony of the operation of a psychic tone in the human digestive tract.

The data just given reveal that the opposed influences of affective states favorable and unfavorable to good digestion are expressed through parasympathetic and sympathetic nerves which are themselves typically antagonistic in their functional relations to the digestive organs.

The Emergency Function of the Sympathico-adrenal System in Strong Emotion. Anger, rage, and fear are the conditions par excellence for studying the bodily reverberations of emotion. The surface manifestations of these major affective states are truly remarkable. They have been studied, observed, and commented upon by a variety of philosophers, psychologists, physiognomists, artists, and natural historians. But the detailed examination of the deep-seated visceral alterations which occur when a strong emotion is in evidence has required the special methods of the physiologist.

As already related, physiological investigations have shown that, when sight and taste and smell of food are attended by agreeable sensations, the cranial autonomic goes into action to promote the processes of digestion and so fortifies the body against times of stress. Evidence has also been presented that interruption of this mild feeling of pleasure by fear or anger is accompanied by sympathetic activity that completely overthrows the quiet conserving processes. The significance of this lies in the fact that the inhibition of digestive functions is merely one of a multitude of effects produced by widespread sympathetic activity at this time. In the angry cat, for example, the hairs of the back and tail are erected, the pupils are dilated, and the eyes become prominent, the blood sugar rises, the heart is accelerated, and widespread vasoconstriction occurs. Likewise, the human being assailed by anger or fear shows a rapid heart, a high arterial pressure, wide pupils, sweating, and high glycemic percentages which lead to the appearance of sugar in the urine (11).

In discussing the organization and physiology of the sympathetic system it was pointed out that conditions which demand an increased outlay of energy are characterized by unit action in this system. The bodily alterations so induced are of the utmost service to the organism. Strong emotion is only one of several conditions attended by these responses. When it is aroused, intense muscular effort is likely to ensue. It is then that the sympathetic most thoroughly fulfils its emergency function. Here it is that domination of the viscera by sympathetic impulses most notably mobilizes the bodily forces for immediate action.

It is a firmly established fact, generally agreed to, that artificial stimulation of the splanchnic nerves, by exciting the secretory fibers distributed to the adrenal medulla, causes a discharge of adrenin into the blood stream (11, 50). Therefore, it might be expected that nerve-impulses are sent out along these fibers and excite the adrenin-secreting cells whenever, under natural conditions, there is general sympathetic activity. Since adrenin has the same actions as sympathetic nerve-impulses, it would, if secreted at such times, cooperate with those impulses to produce widespread effects. Whether this actually occurs or not is a question which was first taken up by Cannon in 1911. The positive results obtained by Cannon and his collaborators from carefully controlled experiments during the last eighteen years have revealed the emergency function of the sympathetic system together with that of its constituent part, the adrenal medulla. The importance of these investigations can be described by saying that they have clearly established the function of an endocrine organ and thrown new light upon the action of the nervous system. As objective studies of emotion they have a special significance for the psychologist.

In this work two experimental methods have been employed to determine the secretory activity of the adrenal medulla in laboratory animals. The first involved the painless taking of samples of blood from the inferior vena cava just above the mouths of the lumbo-adrenal veins by means of a catheter introduced from below through the femoral vein. The adrenin content of caval blood was then tested before and after the experimental procedure by its effect upon rhythmically contracting strips of intestinal muscle which are characteristically inhibited by adrenin in dilutions as great as one part in many millions. Since adrenin is practically the only known substance occurring in blood which has this effect, and since adequate control experiments showed that the adrenal glands are alone able to confer the inhibitory power upon blood, it was considered that the method was reliable (11). The second method of Cannon and his collaborators makes use of the completely denervated heart (12, 18). When the extrinsic nerves of the heart, the vagal inhibitory fibers and the sympathetic accelerators, are cut, the only factors, aside from a rise in blood temperature, which are capable of increasing the heart rate are medulliadrenal secretion, thyroid secretion, and a cardio-accelerating substance which, under certain conditions, may be liberated into the blood from the liver (18). To observe the influence of any one of these humoral stimulating agencies it is necessary only to eliminate the other two. Since all are under nervous control, this may be accomplished by section of the appropriate nerves. The thyroid innervation is interrupted by removal of the stellate ganglia (a part of the usual method of denervating the heart); the innervation of the liver, by severing the hepatic nerves; and the nervous supply to the adrenals, by section of the splanchnics (more conveniently the adrenal factor is eliminated by removal of one gland and denervation of the other). With thyroid and hepatic influences disposed of, the denervated heart becomes a delicate indicator of circulating adrenin. Its reliability in this respect has been abundantly demonstrated.

In 1911 Cannon and de la Paz (11), using the catheter method, first showed that there is an emotional stimulation of medulliadrenal secretion. They noted that, when a cat was frightened (or enraged) by a barking dog, blood drawn from the inferior vena cava just above the opening of the lumbo-adrenal veins repeatedly caused inhibition of the beating intestinal strip, whereas blood removed before the excitement had no effect, and, further, that excitement after removal of the glands did not yield the positive result. In a second report Cannon and Hoskins brought forth evidence that in anaesthetized animals the adrenal medulla is stimulated to secrete by asphyxia and by such sensory stimulation as would have caused very severe pain in a conscious animal. Thus began a long series of experiments in which it was shown that "painful" sensory stimulation and asphyxia, as well as emotional excitement, throw into effective action the secretory nerves of the adrenal medulla. All three of these agencies are closely related in the sense that they are emergency states and produce similar bodily changes.

Among the early investigations leading to the formulation of the emergency theory of sympathetic and adrenal functions were those carried out by Cannon in collaboration with Gray and Mendenhall (11) which showed that injected adrenin, stimulation of the splanchnic nerves, and fear and rage hasten the clotting of blood. The demonstration that the adrenals are necessary for the effect when produced by splanchnic stimulation, and that emotional excitement evokes more rapid clotting only when the splanchnics are intact was further evidence for emotional activation of the adrenal medulla. The proposition that acceleration of clotting is an emergency function of the adrenin secreted during fright and fury needs no argument, for combat with possible injury and hemorrhage is foreshadowed by the circumstances which ordinarily evoke these emotions.

The relative difficulty and intricacy of the catheter method led Cannon to adopt the denervated heart as a preparation for detecting adrenal secretion. The latter is superior to the older method in one important particular, namely, that it tests the adrenin content of blood circulating in the animal under investigation. Since adrenin is an unstable compound that rapidly undergoes inactivation in the blood stream, the denervated heart permits a continuous record showing the duration and subsidence as well as the latency of the secretion. By means of the newer method Cannon in 1919 (12) was able to confirm the results obtained with de la Paz eight years before. In 1925 Cannon and Britton (16) used the denervated heart in connection with a "pseudaffective" state which they were able to produce in cats by ablation of the cerebral cortex and in which are manifested the activities typical of anger. This decorticate sham rage (described in detail in the next section of this chapter) occurs in fits during which there are all the signs of vigorous, widespread sympathetic discharge. In fifteen experiments it was found that the denervated heart had rates averaging 252 beats per minute during the fits of rage and 222 beats per minute during the intervening periods of quiet. After the adrenal glands had been removed, the rate averaged 156 and was increased only four to eight beats during activity. It was concluded that a greatly increased secretion of adrenin is associated with this form of emotional behavior. That this secretion has a notable effect upon the heart, over and above the direct effect of sympathetic nerve-impulses, has been shown by the writer in some recent experiments (not yet published). It was found that when vagal and adrenal influences were excluded, leaving only the accelerator nerves to act, the maximal rates in fifteen experiments averaged 289 beats per minute. Since this is 37 beats over the average of the maximal rates produced by medulliadrenal secretion in the comparable experiments of Cannon and Britton, it is apparent that accelerator impulses are more potent than the humoral agent. But the importance of the latter was demonstrated by the results of an equal number of experiments in which both factors were active (vagi cut). Then the maximal rates attained the extraordinary average of 305 beats per minute. Under the circumstances secreted adrenin obviously reinforces and supplements the direct nervous influence.

The evidence for emotional secretion of adrenin obtained by Cannon and his associates has received, from the very beginning, abundant confirmation. As early as 1912, Elliott (28) reported some observations on the "paradoxical dilatation" of the pupil which occurs after removal of the superior cervical ganglion and is due to increased sensitivity of the radial muscle of the iris to adrenin. He found that this response was evoked in anger, but only so long as the adrenals were intact and innervated. Elliott's findings were later confirmed by Kellaway. Recently Sataké, Watanabé, and Sugawara (49) have studied the rate of medulliadrenal secretion by testing, with the intestinal strip, blood taken directly from the lumbo-adrenal vein. Their dogs were prepared for the experiment in advance by section of the sensory nerves of the loin, thus making it possible to obtain adrenal blood without anaesthesia and without disturbing the animals in the least. It was found that emotional excitement evoked by fastening the animals was accompanied by an increased secretion amounting to several times the rate in quiet existence. Such, in brief, is the direct support given Cannon's contention that the adrenal medulla becomes active in emotion. And from many investigators working in diverse parts of the world and using a variety of methods has come confirmation of his demonstration that other stressful experiences, pain, asphyxia, exposure to cold, hypoglycemia, all produce an augmented secretion of adrenin (11, 22, 14). Contrary to this mass of positive evidence is the doctrine, vigorously put forward by Stewart and Rogoff some years ago, that the output of adrenin is constant and unchangeable (54). This is not the place to weigh and counterweigh the opinions of Stewart and Rogoff. It is sufficient to point out that their theory was based implicitly upon negative results obtained by a method which, in the hands of the only other investigators who have used it, has contributed further evidence in favor of Cannon's view (42, 55). Furthermore, the reader is advised that Cannon and his collaborators (22, 17, 14) have effectively answered, by appeal to experimental evidence, the many theoretical arguments advanced by Stewart and Rogoff in criticism of each and every method that has yielded results contrary to their own.

Most conclusive and most significant of all the evidence for emotional activation of the adrenal medulla is that recently reported by Cannon and Britton (17). They made use of cats in which denervation of the heart and all other operative procedures were carried out by surgical methods. The emotional reactions of these animals were studied later while they were living normally in the laboratory. In such lasting preparations of the denervated heart emotional excitement induced by the presence of a barking dog or by restraint in a holder was invariably accompanied by marked increments in pulse rate. Characteristically they persisted for some time after the stimulus for them had ceased to act; after the cats had been excited by a dog for one minute they were removed from the scene of the disturbance and allowed to lie quietly on a cushion whereupon the faster pulse rate fell off gradually and only attained the quiet rate after 15 or 20 minutes. This result is similar to the prolonged inhibition of gastric secretion noted by Bickel and Sasaki and is related to the common experience that serenity and calmness are not restored for some time after an emotional upset. Control observations were made on the same animals by Cannon and Britton after extirpation of one adrenal and denervation of the other. Although such animals were quite healthy (due to retention of the essential cortex of one gland) and reacted with the same outward vigor to the emotional stimuli, the faster heart rate either wholly failed to occur or was so slight as to be explicable by a rise in temperature.

Of special interest was the finding, in these experiments of Cannon and Britton, that the output of adrenin varies directly with the intensity of the physical effort. For example, the increments in heart rate averaged 22 beats per minute when the excitement was accompanied by slight muscular activity (hissing, snarling, retraction of the ears, twitching of the tail) and 49 beats when it was attended by struggle or aggressive movements of attack. This shows that the bodily processes are more profoundly altered in the full expression of emotion than in the preparatory visceral changes alone. As pointed out by the authors, it follows that "when limitation of the bodily disturbances is desirable, control of the controllable factors, the voluntary muscles, would minimize the effects of an emotional storm." This new evidence is in contradiction to the view that suppression of the outward manifestations of strong emotion may lead to exaggeration of the visceral components of the response. Previous to the work of Cannon and Britton, the observations of Hartman, Waite, and Powell had furnished evidence that adrenal secretion is increased in muscular exercise and is proportional to the extent and duration of the activity.

Gradation of adrenal secretion with the vigor of somatic emotional behavior emphasizes the cooperation of the sympathetic system in activities involving energy changes in the body. The muscular activity which prevails or is likely to prevail in strong emotion is primarily dependent upon an increased oxygen supply. This requirement is met by adaptive changes in the circulatory and respiratory systems. The faster heart and vasoconstriction in the splanchnic region with shifting of blood to other parts are factors which increase the circulation rate and so multiply the number of round trips taken by the oxygen-carrying red blood-corpuscles between the lungs and active tissues (skeletal muscles, central nervous system, and heart muscle itself). These changes are under sympathico-adrenal control. And recent work has revealed another sympathetic effect which is of service in supplying the increased oxygen demand of motion and emotion. Barcroft and his co-workers have proved that conditions of oxygen deficiency—asphyxia, hemorrhage, muscular exercise—cause the spleen to contract, and that when it contracts it discharges into the blood stream an extra supply of red blood-corpuscles. Previously a higher concentration of red corpuscles in the blood after emotional excitement or injection of adrenalin had been observed by Lamson, but he attributed it to occurrences in the liver. Later Izquierdo and Cannon (37) demonstrated a 20-per-cent increase in circulating red cells in cats excited by dogs, have shown that the effect is splenic in origin and is due to a sympathetic discharge to the spleen. As regards the coincident secretion of adrenin, their experiments failed to indicate how great a rôle, if any, it plays in producing this effect. But it can certainly support the sympathetic impulses which act directly upon the heart and splanchnic vessels to bring about the vascular adjustments so necessary to sustained muscular activity. Also, amounts of this hormone corresponding to the amounts secreted have the ability to produce a rapid recovery of the capacity of fatigued muscle to respond to stimulation (11). Its remarkable action in restoring the ability of exhausted animals to continue running has been clearly shown in the work of Campos, Cannon, Lundin, and Walker (9). It was also found by these investigators that cutting the sympathetic nerves to the liver results in a striking reduction of the working ability. They were unable to ascribe this effect to interference with any of the known functions of the liver nerves.

Another aspect of the sympathico-adrenal response in emotion which has a definite emergency value is the increase in blood sugar that it effects by causing a breakdown of the glycogen stores of the liver (glycogenolysis). The utility of such mobilization of carbohydrate lies in the fact that sugar is the optimum source of muscular energy. Since the kidneys do not permit sugar to escape until the percentage in the blood rises considerably above the normal level, the appearance of sugar in the urine, as a consequence of emotional excitement, is an indication of emotional hyperglycemia. Emotional glycosuria in man and animals is an old and frequent observation (11). In 1911 Cannon, Shohl, and Wright (see 11) demonstrated its appearance in cats, found that it varied with the magnitude of the emotional disturbance, and showed that it could not be induced after removal of the adrenals. The latter observation suggested that medulliadrenal secretion is a prime factor in producing the change and that it is of more consequence than a direct sympathetic discharge to the liver. Desiring further evidence on this point, Bulatao and Cannon (8) investigated it in animals manifesting decorticate rage. They found that when the true pseudaffective state developed the blood sugar rose, on the average, to nearly five times its normal percentage. Appropriate operative procedures showed that as a causative factor secreted adrenin is far more potent than impulses over the splanchnic branches to the liver. A recent study by Britton (7) has again demonstrated the prepotency of the humoral factor in throwing into action the glycogenolytic processes. When normal cats were excited by an aggressive dog, there were marked increments in the glycemic percentage. These responses were scarcely modified after section of the sympathetic nerves to the liver. On the other hand, ablation of the medulla of both adrenal glands, although it had no influence upon the general affective response, abolished the rise in blood sugar. It is interesting that a parallelism in the visceral and somatic reactions was again revealed by these experiments. Even so, the release of considerable quantities of sugar into the blood stream when emotion is accompanied by only slight muscular activity seems to be in anticipation of a greater demand.

The experimental evidence presented in the foregoing paragraphs gives reality to the statement that in emotion the viscera are dominated by activity in the sympathetic division of the autonomic system. In the production of these changes sympathetic impulses and adrenin, itself secreted as a result of sympathetic discharge, cooperate. One change, the speeding of blood-clotting, is solely the result of increased adrenal secretion. Another, glycogenolysis, appears to be chiefly, if not entirely, due to the humoral factor. In the case of cardiac acceleration, the available evidence suggests that the direct neural influence is somewhat prepotent, the humoral important but supplementary; the same is probably true of splenic contraction and splanchnic vasoconstriction. It was previously thought that adrenin has an effect apart from sympathetic impulses in increasing the excitability of fatigued skeletal muscle, but if the testimony of Orbeli (see 30) as to the beneficial influence of sympathetic impulses on fatigued muscle should be confirmed, it would suggest that here, too, the neural agency may be of importance.

The Central Mechanisms for Emotional Reactions

At various times "the seat of the emotions" and the central mechanisms controlling emotional behavior have been sought at the highest level of the central nervous system, namely, the cerebral cortex. But however reasonable the assumption may be that the neural processes underlying emotional consciousness are cortical, it does not follow that the bodily changes which make up emotional behavior are due to a discharge of cortical origin. In fact certain considerations and many experimental facts indicate that the bodily responses typical of one strong emotion, rage, are operated by the more primitive parts of the brain.

First of all, there is the fundamental nature of the behavior attending this emotion. It is a behavior which is called forth by the urgency of certain definite circumstances and it is directed toward the preservation of the individual. It constitutes a reaction which is primitive, energetically purposive, and common to many different members of the vertebrate series. In short, it possesses many features in common with such elementary responses as sneezing, coughing, and sucking. An illuminating summary of the similarity of the rage-reaction to these simple reflexes has been given by Cannon (15): "First, its occurrence in the early months of even so highly developed an organism as the human infant indicates that its neural pattern, like that of the reflexes mentioned above, is congenitally inwrought in the central nervous apparatus. Second, as in the reflexes, it is a prompt response to the appropriate stimulus. Again, it is a constant and uniform response—so much is this so, indeed, that there is no mistaking its character, whether it be manifested by the diverse races of man or by the lower animals. It is like the reflexes, also, in being a permanent mode of reaction; throughout life the characteristic display of the rage-response may be suddenly evoked in all its elaborateness. Further, it is a response to a fairly definite stimulus—an inner stimulus which arises when there is a hampering or checking of motion or an opposition to one or another primary impulse. Finally, the rage-response is like the simple reflexes in being useful." Obviously the rage-response has all the earmarks of a subcortical reaction. Its prompt, uniform, and stereotyped character places it in sharp contrast to the type of behavior subserved by the cortex. The latter is characteristically complex, uncertain, temporary and modifiable by experience. The behavior of an enraged cat is no more complex and no more the result of experience than are its postural and righting reflexes, which, as shown by the late Professor Magnus of Utrecht (45), are managed in an automatic inborn manner by neural mechanisms which lie wholly in the brain-stem and spinal cord. As regards their degree of complexity and integration and their immunity to modification, there is every similarity between the reactions of an angry cat and those reactions of a falling cat which land it invariably upon all four feet.

The foregoing considerations certainly suggest that the expression of this emotion is dependent upon the older, more primitive divisions of the central nervous system. There is much experimental evidence to show that this is actually the case.

When a dog or cat is deprived of its cerebral hemispheres the capacity for the exhibition of anger is by no means abolished. After complete removal or disconnection of all parts of the cerebral cortex and with some injury to corpora striata and the dorsal part of the thalamus, Goltz's dog (33) lived in good health for more than 18 months. The animal was reduced to a state of idiocy; it lost all of its learned behavior, showed no ability to regain the loss, and did not acquire any new modes of response. Its emotional activity was confined to a reaction, capable of regular elicitation, which closely resembled the picture of rage as seen in a normal dog. The same was true of the similar dog described by Rothmann (48). Dusser de Barenne (27) prepared and studied for months two decorticate cats which showed marked emotional responses to various disturbances, innocuous as well as mildly harmful; merely lifting the animals caused energetic movements of defense and those reactions so characteristic of the angry cat—spitting, growling, and erection of the hair. The trivial and often irrelevant nature of the conditions which would evoke the rage-reaction in these four animals is noteworthy. It was readily induced in Goltz's dog by merely pinching the skin. Taking the animal from its cage invariably caused barking, growling, biting, and every evidence of violent protest in spite of the fact that this procedure was the usual sign for feeding and would have become quite agreeable to any normal dog. Snarling and growling were obtained in Rothmann's dog by gentle scratching of its back, and the presence of a fly on the creature's nose sent it into a fit of rage. It is disappointing that the reports of the emotional reactions of these animals mention only one symptom of sympathetic activity, the erection of hair in the case of the cats. Other superficial signs of sympathetic discharge were undoubtedly present, but were evidently overlooked. A short time ago, through the kindness of Dr. Stanley Cobb, the writer was afforded the opportunity of observing the behavior of a surviving decorticate cat, which closely resembled Dusser de Barenne's animals. It exhibited well-marked signs of anger when disturbed. Light pinching of its tail regularly evoked hissing and biting with lowering of the head and arching of the back. Accompanying these somatic activities were the surface signs of widespread sympathetic discharge—erection of the hair of back and tail, retraction of the nictitating membranes, dilatation of the pupils, and separation of the lids (exophthalmos).

The observations of Goltz, Rothmann, and Dusser de Barenne, showing that cortical ablation renders the mammal excessively prone to the display of anger, led Cannon and Britton (16) to make use of decorticate cats in the acute condition, i.e., animals which do not recover from the effects of the operation, for the study of emotional activation of the sympathetic. It was found that after disconnecting the cortex from the brain-stem there appeared upon removing the anaesthetic "a group of remarkable activities such as are usually associated with emotional excitement—a sort of sham rage." They included: lashing of the tail; arching of the trunk; thrusting and jerking of the limbs in the thongs which fastened them to the animal board, together with protrusion of the claws and clawing movements; snarling; movements of the head from side to side with attempts to bite; very rapid panting with mouth open and movements of the tongue to and fro. These somatic activities were accompanied by signs of a vigorous sympathetic discharge; erection of the tail hairs; sweating of the toe-pads; dilatation of the pupils; large increments in heart rate and arterial pressure; an abundant secretion of adrenin; and, as later shown by Bulatao and Cannon (8), an increase in blood sugar up to five times the normal concentration. These activities occurred without special stimulation in "fits" lasting from a few seconds to many minutes with intervening periods of quiet during which a fit could be evoked by the slightest disturbance of the animal.

The investigations to which reference has just been made have established the fact that the decorticate cat or dog, in the chronic as well as the acute condition, is capable of displaying a type of behavior which closely resembles that of the infuriated normal animal. They give strong support to the view that the expression of rage is not dependent upon cortical activity, but is the result of a motor discharge of subcortical origin. In addition they give some idea of the character of the central processes underlying the normal production of the rage-reaction. In 1884 the English neurologist, Hughlings Jackson, made a suggestion which has greatly influenced modern studies of the nervous system and which has received much support in recent years, especially in the writings of Head (34, 35). It is to the effect that the nervous system is organized in such a way that the primitive reactions of the older parts are prevented from playing a dominant rôle in behavior by the inhibitory influence of higher levels. Thus the cerebral cortex normally holds in check those activities of the lower, more archaic centers which would seriously interfere with its more discriminative reactions. The expression of emotional excitement is just this sort of activity. On the basis of the doctrine of Jackson and Head its occurrence after removal of the cortex may be regarded as a "release phenomenon" due to the freeing of a subcortical mechanism from cortical restraint. In agreement with this interpretation is the excessiveness and easy elicitation of the decorticate rage, qualities which suggest hyperexcitability of the centers involved. Its appearance in long-surviving decorticate animals precludes its being the result of irritation incident to the operation. When an outburst of rage takes place in normal life, we may suppose that it is engendered either by a diminution of the cortical inhibition or by a subcortical excitation of sufficient magnitude to offset the cortical control.

Some idea of the subcortical region responsible for the emotional activity of decorticate animals can be gained from experiments in which cerebral ablation has not been followed by such behavior. Since it was first described by Sherrington (52), the decerebrate animal, in which the brainstem is truncated at the level of the mid-brain (Figure 3), has been widely employed in the analysis of reflex action. The testimony of all who have investigated such preparations is consistent in showing that they never exhibit the typical decorticate rage. To be sure, Woodworth and Sherrington (59) have described certain responses, expressive of affective states, which they were able to evoke in decerebrate cats by strong stimulation of sensory nerves; and similar "pseudaffective reflexes" were noted by Bazett and Penfield (4) in the decerebrate cats which they were able to keep alive and study for periods of time varying from several days to

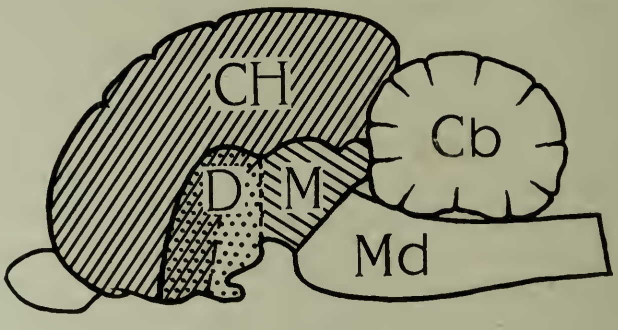


FIGURE 3

Median Section of the Mammalian Brain

CH, cerebral hemispheres; *D,* diencephalon (indicated by dots); *M,* mesencephalon (mid-brain); *Md,* medulla oblongata; *Cb,* cerebellum. The cross-hatching, from right downward to left, marks the portion of the brain which can be removed without interfering with the expression of rage. (From: Cannon, W. B. The mechanism of emotional disturbance of bodily function. *New England J. Med.,* 1926, **198**, 877-884.)

three weeks. It is, however, important to distinguish between the pseudaffective activity of these mid-brain preparations and the sham rage of decorticate animals. The former, as exemplified by the experiments of Bazett and Penfield, consisted of such isolated items of behavior as biting, clawing, and waving of the tail; it never attained a general affective state and was typically brought forth by a stimulus connected with some habitual mode of response. And although in the experiments of Woodworth and Sherrington these reflexes, evoked by strong afferent stimulation, had a certain "width of coordination," it was pointed out that "they never amounted to an effective action of attack or escape." On the other hand, the sham rage of the decorticate animal in the chronic as well as the acute condition is elicited by the slightest disturbance of any kind, is astonishingly intense, and possesses a width and energy of expression that makes it unmistakably the counterpart of intense fury in the normal animal. This more general, more energetic, and more easily evoked behavior must be dependent upon neural mechanisms which lie below the cortex and above the mid-brain.

The central source of the characteristic sympathetic activity of strong emotional excitement is a matter of primary importance in any consideration of the neural basis of emotional expression. A central representation of certain fractions of the sympathetic system can be discerned in the medulla oblongata. The bulbar vasoconstrictor center, the bulbar cardio-accelerator mechanism, and the center for the reflex secretion of adrenin located by Cannon and Rapport (23) in the upper part of the medulla may be regarded as sympathetic centers. These and other bulbar mechanisms are necessary for the reflex activation of definite groups of sympathetic fibers and are responsible for tonic discharges along sympathetic paths. They still perform such duties when separated from the higher parts of the nervous system, but the absence of generalized sympathetic activity after decerebration and its presence in the sham rage of decorticate animals indicate that these individual bulbar centers are not sufficient for the emotional activation of this system. In fact, the marked tendency of the sympathetic to discharge as a whole under conditions of stress, and particularly during strong emotion, implies the existence of a higher dominant central mechanism which when active will produce activity in the entire thoraco-lumbar outflow.

With these points in mind, Bard (3) undertook an investigation directed toward the delimitation of the part of the brain-stem responsible for the sham rage of the decorticate cat. It was felt that such a study would further elucidate the neural basis of this emotion. Also, there was every indication that the region sought represents the dominant center of the sympathetic system. The experimental procedure adopted was that of ablation of varying amounts of the brain-stem after removal of the hemispheres. It was found in a series of about fifty acute experiments that the typical sham rage, accompanied by vigorous discharge of sympathetic impulses, regularly occurred after removal of hemispheres, corpora striata, and the cranial half of the diencephalon (Figure 3). It also appeared when, in addition, much of the dorsal part of the diencephalon, the thalamus proper, was cut away. But it invariably failed to develop after transecting the caudal extremity of the diencephalon or the cranial portion of the mid-brain. From these observations, it was concluded that the discharge of nervous impulses which evokes the extraordinary emotional display of the decorticate animal is conditioned by central mechanisms which are situated in the more ventral part of the caudal half of the diencephalon, probably in the hypothalamus. A part of the central arrangements for the expression of rage are of course located below this level. Certain individual elements of the entire reaction may be induced in the spinal animal, and still more in the bulbospinal or mid-brain preparations, but the results of this investigation indicate that it is only when the diencephalic mechanism is present that these elements can be readily welded together to form the rage-reaction.

This same investigation shows that at the level of the diencephalon the visceral activities of rage are linked and coordinated with the somatic. The two were found to be inseparable. It is significant, in this connection, that considerable physiological evidence has been accumulated to show that there is a diencephalic representation of the sympathetic. Karplus and Kreidl (40), for example, have observed that local stimulation of the hypothalamus in narcotized animals causes various sympathetic discharges, and Houssay and Molinelli (36) reported that similar stimulation induces medulliadrenal secretion. In the light of the interpretation given above, these effects should be regarded as the results of stimulation of the dominant center of the sympathetic.

On the basis of the foregoing account it may be stated with some assurance that the nervous organization for the display of rage, both in its somatic and its visceral aspects, is located in the base of the diencephalon, a phylogenetically ancient part of the brain, common to all members of the vertebrate series. So much is true of rage. What of other primitive emotions? Have they too a neural basis in the older parts of the nervous system?

Certain it is that the expressions of fear, joy, and grief are typical inborn reaction patterns. They are prompt, uniform reactions, similar in many different species, present at birth or emergent as permanent unconditioned responses later (a point grossly disregarded in some of the child studies of the behaviorists). Like the expression of rage they are allied to the simpler reflex figures and are thus distinguished from the reactions of the cortex which are adjustments and modifications of behavior to suit the ever changing events of the outer world. We should expect to find that the essential neural basis of these reactions, like that of rage, lies in the brain-stem. Little evidence that this is the case comes from animal experimentation. Many years ago Bekhterev (5) reported signs of pleasure in freshly decorticated animals, purring in cats and tail-wagging in dogs, and he stated that these responses disappeared after removing the thalamus. But Bekhterev's is an isolated experience which has never been shared by other workers who have dealt with decorticate animals. On no occasion did Goltz's dog give any evidence of joy or pleasure nor did she show the slightest sign of sexual emotion. The same was true of the other hemisphereless animals mentioned above. Possibly the rage-reaction, so prominent in such animals, dominates their behavior to the exclusion of other emotional reactions of which they may well be capable. Perhaps the latter could be revealed by appropriate operative procedures. At any rate, to conclude that they are not subcortical reactions would be premature and out of harmony with the evidence about to be presented.

When we turn to the experience of the clinic, we find a mass of evidence to show that there is subcortical management of several types of emotional expression. This has recently been presented in some detail by Cannon (13, 15). It consists of examples of emotional overactivity and of emotional paralysis. In the former there are generally good reasons to suspect removal of cortical control; in the latter there are usually signs of subcortical impairment of function.

First of all are certain phenomena associated with anaesthesia. The general anaesthetics, such as ether, chloroform, and nitrous oxide, together with alcohol, are drugs which act as depressants, not as excitants (24). All parts of the central nervous system, however, are not equally sensitive to their action. Indeed they are useful in surgery because they first depress the cortical processes associated with consciousness, next reflexes in general, and, last of all, the bulbar centers concerned with respiration and other essential functions. Accordingly they may be used to produce insensibility and muscular relaxation without endangering life. When any one of these agents is administered the stage of relatively deep surgical anaesthesia is preceded by a primary stage of imperfect consciousness and a secondary stage of excitement. During the excitement stage there may be the most remarkable emotional display; the patient is likely to weep and groan, sing joyously, laugh uproariously, or exhibit all the aggressive signs of rage. While such emotional behavior is going on, the surgeon may—as is sometimes necessary—perform an operation of considerable extent and gravity, and yet later the patient, when wholly conscious, will testify that he was not at all aware of what was happening. It is clear that in such cases vigorous emotional behavior occurs in the absence of emotional consciousness. This means that although the cerebral cortex is narcotized and inactive, lower centers, themselves not depressed, act in much the same way as after surgical decortication. The anaesthetic functionally decorticates, lessens or destroys for a time cortical inhibition of lower centers, and so releases them to carry out their characteristic motor activities.

In further support of a subcortical origin of various kinds of emotional behavior in human beings are many findings of clinical neurology. When a destructive lesion interrupts the voluntary motor path from the cortex, paralysis of willed movement may not be attended by any impairment of emotional expression. The deficiency is most often unilateral, but may be bilateral. Such patients are incapable of moving one or both sides of the face, but when a sad or joyous situation develops, the muscles of facial expression go into action to give both sides of the face an expression of sadness or gaiety (57, 58). In single or double hemiplegia with partial or complete facial paralysis or, more commonly, in "pseudobulbar palsy," a condition involving volitional facial weakness, there may be fits of exaggerated, forced, uncontrollable laughing or crying brought on inevitably and with extreme facility by almost any stimulus. A number of such cases have been described and discussed at length by Kinnier Wilson (57, 58). Two will serve as typical examples. One, a young pseudobulbar patient, had suffered a complete loss of all voluntary facial movement so that he could not voluntarily close his eyes, elevate and retract the corners of his mouth, close or open his mouth, bite, swallow, utter a vocal sound or cough. In contrast to this paralysis of voluntary movement was the preservation of emotional movements of laughing; he went off into rounds of laughter on the slightest provocation so that his existence was one long roar of laughter. Another of Wilson's patients, a woman, became peculiarly lachrymose after a stroke had produced a right hemiplegia with aphasia; the least emotional stimulus threw her into a fit of crying with copious tears so that bout after bout of weeping succeeded one another through the day. In such cases as these the over-reaction may be confined to either laughing or crying, or both modes of expression may occur at different times in the same individual. Uncontrollable pathological laughing or crying are typically allowed by lesions of the voluntary paths from the cortex to the medulla and spinal cord, and Wilson states that "the more severe the volitional facio-respiratory paralysis, the more exaggerated is involuntary innervation of the same mechanism." This fact suggests that interruption of the voluntary cortical control leads to unbalanced activity in a separate neural apparatus which is the source of the emotional behavior and which under normal conditions is held in check by cortical activity.

Because the lesions associated with pathological laughing and crying most often involve cortico-thalamic fibers it has been supposed that the abnormality is the result of freeing the thalamus from a cortical check. Brissaud long ago asserted that integrity of the thalamus is essential for the appearance of spasmodic uncontrolled laughter and weeping, but Wilson (57, 58) objects to this view for the reason that they may occur when the thalamus itself is grossly diseased. Nevertheless, there is evidence to show that when the base of the thalamic region, the part essential for the rage-reaction, is the site of a morbid condition there is typically a loss of emotional expression. This fact is mentioned by Wilson, but it is interpreted by him as due to the interruption of fibers of cortical origin which he believes are "the routes taken by emotional impulses to modify the facio-respiratory synkinesis in the direction either of laughter or the reverse." For the existence of such paths in the brain-stem, especially in the mid-brain, Wilson presents good evidence, both clinical and physiological, but his assumption that they are of cortical origin is not convincing, for, as he himself admits, no case of emotional paralysis originating in a cortical lesion has yet been recorded. On the other hand, lesions which separate the cortex from lower levels are commonly followed by exaggerated emotional activity. It is more in accord with the evidence at hand to think in terms of a subcortical organization for emotional expression located or centered in the under part of the thalamus, connected by efferent paths with the primary motor nuclei of medulla and spinal cord, and normally subject to cortical restraint.

The occurrence of emotional paralysis with retention of volitional normality is another datum of clinical experience which bears upon the present discussion. There are patients who can voluntarily move both sides of the face in a normal symmetrical manner, yet one side remains motionless whenever they laugh in fun, weep in sorrow, or make a grimace in pain. The lesions provocative of this effect are, as far as the writer has been able to determine, invariably subcortical. Kirilzev (41) has described a man with emotional paralysis of the right side of the face in whom autopsy disclosed a tumor in the center of the left thalamus. Wilson (57, 58) reports three similar cases; in one there was a tumor of the subthalamic region on one side; and in the other two there were lesions in the mid-brain. More complete loss of emotional expression is sometimes seen when the ventral part of the diencephalon is involved in a pathological process. Fulton and Bailey (31) have pointed out that among the symptoms of tumors of this region is a curious emotional negativism; for example, one of their patients appeared to be devoid of all emotional expression, exhibited a "fatuous serenity of mind with complete failure to appreciate the gravity of his own physical condition." Also, in cases of "narcolepsy" with disease of the region of the third ventricle the expression and the feeling of emotion may be quite absent; such patients meet jibes and insults with utter indifference and give no evidence of affective response when some tragic happening is called to mind. All these instances of impairment of emotional expression can be reasonably accounted for on the basis of a defect in the subcortical mechanism suggested above.

The foregoing observations, physiological, clinical, and pathological, point to the diencephalon as the region in which resides the neural organization for the expression of various emotional states. It also follows from the same evidence that inhibitory cortical influences normally prevent the primitive activities from dominating behavior. The subcortical processes are at all times ready to seize control of the motor reactions and when the cortical check is released they do so promptly and with elemental vigor. Thus while the behavior attending the elemental emotions is certainly not due to a nervous discharge of cortical origin, we must ascribe to the cortex an important rôle in the production of such behavior. As an efferent projection system the cerebral cortex has a negative function in regard to these activities. Furthermore, as the neural basis for the conditioned responses of the higher mammals (47), the cortex greatly increases the number of circumstances which are capable of acting as emotional stimuli. By means of the cortex the inborn, stereotyped, emotional reactions, like the secretion of saliva and other simple reflexes, become, as the result of experience, conditioned responses. When conditioned stimuli evoke emotional responses we must suppose, on the basis of the argument just developed, that they do so by releasing the cortical check.

Theory of Emotion

The term emotion, as it is reasonably employed, implies two things: a way of acting and a way of feeling. The foregoing account has been limited to a consideration of the nature and the neural basis of the bodily changes in emotion. It would, however, be a truncated and incomplete treatment of the subject that did not mention the relation of these physiological phenomena to emotion as a mode of feeling. The nature of that relation constitutes a problem of psychological importance, and it may be approached with advantage from the point of view of this chapter.

A cross-section of contemporary psychological literature reveals a fairly widespread tendency to accept as a solution of this problem the familiar theory of the emotions which is associated with the names of William James and C. G. Lange.

The James-Lange Theory. Broadly stated, this theory holds that consciousness of the bodily disturbances is the essential element in emotional consciousness. Contrary to the popular conception which regards emotional display as the sequence of emotional consciousness, it insists that the bodily manifestations of emotion must be interposed between the perception of the exciting fact and the occurrence of the mental state termed emotional. Expressed in terms of neural happenings, the theory is that the emotional stimulus (an object or a thought) evokes a motor discharge to effector organs whereupon afferent impulses course back from the reacting organs and throw into action the cerebral processes which underlie emotional feeling. Such is the core of the hypothesis put forward by James, first in 1884 (39), and re-explained by him ten years later (38). As regards the source of the afferent impulses which transform the "object-simply-apprehended" to the "object-emotionally-felt," James assumed that they arise generally from somatic and visceral effectors. Always, however, he was inclined to attribute greater importance to the backflow from the viscera. Lange, who in 1885 independently announced a similar theory, elaborated it almost entirely on the basis of changes in the circulatory system. According to him (39, p. 73) stimulation of the vasomotor center by whatever arouses the emotions is the "root of the causes of the affections, however else they may be constituted, and is fundamental to the physiological phenomena which are the essential components of the affections." Common to both James and Lange was the tendency to limit their theory to the explanation of the "coarser" emotions, fear, anger, joy, and sorrow.

The arguments advanced by the original sponsors of the theory were not based on any crucial experimental tests. Lange enumerated the symptoms of the major emotions, claimed that they are all referable to vascular changes and distinguished between one emotion and another on the basis of different bodily accompaniments. He pointed out that physiological effects attended by emotion may be induced by a variety of causes (wine, hashish, certain mushrooms, opium, a cold shower) which, he thought, are "utterly independent of disturbances of the mind." In addition he argued (39, p. 66) that abstraction of the bodily symptoms from a frightened individual leaves nothing of his fear—"let his pulse beat calmly, his look be firm, his color normal, his movements quick and sure, his speech strong, his thought clear; and what remains of his fear?" This same speculative argument was one of the chief lines of evidence cited by James in support of his contention. He urged (39, p. 102) as the vital point of his whole theory this introspective consideration: *"If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, we find we have nothing left behind,* no 'mind-stuff' out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains." In further support of his theory James mentioned certain experiences in which bodily effects appear to be antecedent to the arousal of an emotion or an emotional idea; he summoned the testimony of fellow psychologists and of actors that sometimes, at least, voluntary mimicry of emotional behavior is attended by the appropriate affect.

Experimental and Observational Tests of the James-Lange Theory. Since the time when the James-Lange theory was developed, physiological experimentation and neurological observation have provided a number of objective facts which bear directly upon it and give considerable insight as to its validity. These we shall consider.

1) First of all, there are the numerous events occurring in the viscera during great emotional excitement. According to James, they play a major rôle in giving to consciousness an emotional hue. According to Lange, emotion is wholly the product of vascular changes. Pertinent to these assumptions are the following experimental facts.

When the visceral processes which typically occur in emotion cannot occur, emotional behavior is not altered. The first proof of this was obtained by Sherrington (52) who sectioned the cervical spinal cord and vagus nerves in a number of dogs, thereby destroying all connection of the brain with the viscera of abdomen and thorax and separating the vasomotor center from all vasomotor nerves except the tiny vasodilator components of the seventh and ninth cranial pairs. As previously mentioned, Cannon, Lewis, and Britton (19) have kept cats in a healthy state for many months after removal of the entire sympathetic division of the autonomic system, the division which in fear and rage operates to produce the visceral changes typical of those emotions. These operations, which removed surgically the sensations from the viscera to which Lange attributed the whole and James the major part of the felt emotion, had no effect upon the emotional responses of the parts which remained capable of reacting. Cannon reports (13) that in the sympathectomized cats "all superficial signs of rage were manifested in the presence of a barking dog—hissing, growling, retraction of the ears, showing of the teeth, lifting of the paw to strike—*except* erection of the hairs." In Sherrington's dogs, reduction of the field of perception to the head, neck, and a small portion of the fore limbs with practically total visceral paralysis and anaesthesia, produced no lessening of their emotional character. In anger, joy, disgust, and fear the innervated effectors reacted as before. These experiments demonstrate clearly that if the operated animals experienced any emotional feeling, that feeling was not subsequent to or in any way dependent upon visceral changes. To that extent they contradict the James-Lange theory. It has been argued, however, that such experiments have little or no bearing upon the theory because the animals' consciousness may have been modified under the experimental conditions, that in fact the marked emotional display might not have been accompanied by any emotional feeling. But, as Sherrington has remarked, it is difficult to think that the perception initiating the wrathful expression should bring in sequel angry conduct and yet have been impotent to produce "angry feeling." It may be added here, in anticipation of later consideration, that there have been human cases (25) in which elimination of the entire sympathetic system (and all of the somatic system below the neck as well) failed in any way to alter the emotional *feelings* of the patient.

If the emotions are the consequences of the visceral aspects of emotional behavior, we should expect them to follow these changes whenever and however they are produced. But this is not the case. We have already seen that secretion or injection of adrenin evokes all the visceral changes characteristic of and common to fear, rage, and other strong emotional excitement. But Marañon and others (see 13) have reported that injections of this substance into normal human beings in amounts sufficient to evoke these changes did not produce an emotional experience. They merely gave rise to coldly perceived sensations of palpitation, of diffuse arterial throbbing, of oppression in the chest, of trembling, of chilliness, of nervousness. In certain cases these sensations were coldly reminiscent of previous emotional experiences in which such changes had occurred; the subjects described their feelings by such remarks as, "I feel as if afraid," "as if moved," "as if I had a great fright yet am calm."

Pertinent in this same connection is Cannon's criticism (13) of the James-Lange theory that the same visceral changes occur in such different emotional states as fear and rage and in such non-emotional states as exposure to cold, hypoglycemia, asphyxia, and strong muscular exercise. All of these states rouse the entire sympathetic system to activity and so affect in a stereotyped way the viscera which are under sympathico-adrenal influence. As Cannon states, the responses in the viscera are too uniform to offer a satisfactory means of distinguishing states which, in man at least, are very different in subjective quality. Physiological evidence thus completely refutes the assumption made by Lange (39, p. 62) that the difference between the different emotions is to be explained by a difference in vasomotor reactions. James, at least in the final expression of his views (38), did not expressly attempt to differentiate emotions on the basis of different bodily changes, and it would not be fair to employ this argument against his theory. He did, however, greatly emphasize the importance of visceral factors in the total reaction, which, he claimed, is the essential cause of any affective state. For example, when it was objected that laughing from tickling and shivering from cold, give rise to mere local bodily perceptions, not to real mirth or an emotion of fright, he answered that in no such instance is "the reproduction of an emotional diffusive wave complete. Visceral factors, hard to localize, are left out; and these seem to be the most essential ones of all. I have said that where they also from any inward cause are added, we have the emotion; and then the subject is seized with objectless or pathological dread, grief, or rage as the case may be." In complete refutation of this is the demonstration (21) that with shivering from cold alone there occur the very same visceral changes as occur in actual fright, and that the more intense the shivering the more intense the visceral changes. In this non-emotional state and in others (e.g., running) the *total reaction,* including the visceral reaction, is the same as in fright and yet there is a notable absence of the emotion which the theory of James demands.

2) Supporters of James's theory cannot escape the logic of the evidence that visceral factors are inadequate to explain the genesis of emotion by affirming that sensations of the positions and tensions and movements of skeletal muscle constitute the felt emotion and vary to make the differentia of emotion. The following facts will show how untenable is such an assumption.

In those cases of pathological laughing and crying described by Kinnier Wilson (57, 58) and already mentioned in this chapter, the apparent, visible emotion by no means always corresponds to the patient's real feelings. Indeed these patients, whose emotional feelings are quire normal, suffer greatly in mind from the contradictory expression of their real feelings. Laughing uproariously they may feel sad, weeping profusely they may feel hilarious. Such cases are numerous, interesting, and convincing. "With all the outward appearances of mirth and hilariousness, and with concomitant activity of visceral mechanisms," writes Wilson, "the individual may not only not feel happy, but his state of mind may be in patent conflict with the apparent emotion." "It is clear," Wilson continues, "that the James-Lange hypothesis must be materially modified if it is to be brought into line with observations such as have here been recorded, with no complete fusion between peripheral and cerebral components."

Other clinical evidence of a lack of parallelism between psychical and somatic elements in emotion is available. Wilson has pointed out the frequency of cases of facial emotional paralysis in which the patient readily feels and is acutely conscious of experiencing normal emotional states in response to the usual stimuli. There may be a mask-like expression behind which is experienced the full play of the emotions. A patient thus suffering from complete absence of facial expressional movement "was very sensitive on this point, and termed it his greatest misfortune that he was forced to be joyful or sad without making any demonstration to his fellow-creatures." But still more significant is the presence of normal emotional feelings in patients suffering from total or nearly total immobility of the skeletal musculature. Such is the condition in advanced stages of paralysis agitans with rigidity, in terminal arthritis deformans, in the terminal stages of tabes, and in certain forms of progressive muscular atrophy. Dana (25) remarks that in his experience such patients have normal subjective emotional reactions. This neurologist also reports the case of a woman who, as the result of a fracture of the neck, suffered complete paralysis of the skeletal musculature of trunk and all four extremities with complete loss of cutaneous and deep sensibility from the neck down. She lived for nearly a year without change in personality or character, showed and felt emotions of grief, joy, displeasure, and affection. It may be added that the possibility of an emotional discharge of sympathetic impulses was eliminated. Such cases are not uncommon. We may say with Dana that it is difficult to understand, on the theory of James, why there is no change in them emotionally.

The clinical studies of Head (34) have thrown much light upon the conditions underlying the production of feeling tone. He has described and studied many cases of unilateral lesions of the thalamus, in which the characteristic symptom is "a tendency to react excessively to all potentially affective stimuli." The thalamus, it must be understood, is the great sensory ganglion of the brain-stem; it receives and forwards all sensory impulses (except olfactory) destined for the cerebral cortex. Thalamic lesions may therefore cause deficiencies in sensation. But they also have the effect just noted. This is explained by Head as due to the fact—and there is adequate evidence for it—that a part of the thalamus, "the essential center of the thalamus," has the ability to add to affective experience, that this function is normally under the inhibitory influence of the cortex and, when released from cortical control by a lesion which interrupts cortico-thalamic fibers, induces an excessive affective response to certain stimuli. Under such circumstances, pin-pricks, painful pressure, immoderate heat or cold, all produce far keener discomfort when applied to the body on the side of the lesion; also agreeable stimuli, such as moderate warmth, may evoke intense pleasure when applied to the damaged side. Again, these cases have an excessive sensory reaction to ordinary emotional stimuli. A woman, very appreciative of certain kinds of music before her illness, found that while ordinary sounds left her cold any music capable of stirring her emotions induced disturbing sensations over the afflicted side of the body. Another patient had to keep away from church because the singing of hymns evoked unbearable agony on the damaged side of the body. These cases show that the feeling-tone of any sensation is a product of thalamic activity. Now it is highly significant, as Cannon (13) has pointed out, that in these patients "sensations which underlie the appreciation of posture are entirely lacking in feeling-tone." Hence, the afferent impulses from skeletal muscles which adherents of James's theory have relied upon to provide the extra-visceral part of felt emotion are the very ones which lack the quality to serve the purpose.

The real and essential quality of an emotion is to be found neither in sensory returns from the viscera nor in those from skeletal muscle. Furthermore, the fact that in non-emotional states (exposure to cold, muscular exercise) bodily activity as a whole is the same as in states rich in emotional feeling (fright, anger) makes it very difficult indeed to attribute the emotion to the total reaction. The theory of a peripheral source of emotional experience is thus opposed by overwhelming experimental evidence. The most that can be said for it is that bodily reverberations may add slightly to, and so reinforce, the central processes which underlie emotional consciousness, and that visceral changes may, by a process of association, lead to a calm, non-emotional recall of previous emotional experiences.

A Theory of Emotion Based on Diencephalic Processes. As we have seen, every relevant experimental fact points away from the periphery and directly toward the brain as the site of the processes which determine whether or not a stimulus shall give rise to emotional feeling. What is the nature and the locus of these all-important central processes?

Dana (25) and Cannon (13) have each proposed the theory that emotion results from the action and reaction of the cerebral cortex and the diencephalon. This theory, unlike the James-Lange theory, has much experimental support. It is based upon a number of established facts. First of all, there is the evidence that at the base of the thalamic region are located the neural mechanisms responsible for emotional behavior, mechanisms capable of independent discharge but normally held in check by the cortex. At the same time the cortex is the immediate site of emotional consciousness. But we know that thalamic processes are a source of affective experience, that bodily sensations, such as are sometimes associated with emotion, may be thalamic in origin. Accordingly this theory proposes that at the same time that the diencephalon discharges the motor impulses which produce the emotional behavior it discharges upward to the cortex afferent impulses which throw into action the cortical processes which underlie emotional consciousness. As stated by Cannon, the essence of this theory is that *"the peculiar quality of the emotion is added to simple sensation when the thalamic processes are aroused."*

The neural events which, on this new theory, occur in emotion are described by Cannon as follows: "An external situation stimulates receptors and the consequent excitation starts impulses towards the cortex. Arrival of the impulses in the cortex is associated with conditioned processes which determine the direction of the response. Either because the response is initiated in a certain mode or figure and the cortical neurones therefore stimulate the thalamic processes, or because on their centripetal course the impulses from the receptors excite thalamic processes, they are roused and ready for discharge. That the thalamic neurones act in a special combination in a given emotional expression is proved by the reaction patterns typical of the several affective states. These neurones do not require detailed innervation from above in order to be driven into action. Being *released* for action is a primary condition for their service to the body—they then discharge precipitately and intensely. Within and near the thalamus the neurones concerned in an emotional expression lie close to the relay in the sensory path from periphery to cortex. We may assume that when these neurones discharge in a particular combination, they not only innervate muscles and viscera but also excite afferent paths to the cortex by direct connection or by irradiation."

Cannon (13) has pointed out how well the new theory fits all the known facts. It accounts for the coincidence of the bodily changes and the emotional experience and so explains how James and Lange could reasonably make the suggestion which they did. It explains how the assumption of an attitude does sometimes *help* to establish the emotional state which the attitude expresses, for then the normal cortical inhibition of the lower neurons with reference to that attitude is abolished so that they are already released when a real emotional stimulus acts. On this theory are readily explained the cases of unilateral release of the thalamus from cortical control with accompanying ipsilateral intensification of emotional tone. But, as Cannon has urged, such cases present an insurmountable obstacle to the James-Lange theory, for the viscera cannot function by halves, these patients do not engage in one-sided expression of their emotions, and the impulses sent back from the disturbed periphery are bilateral; the unsymmetrical feeling is due to the organ which is functioning unsymmetrically—the diencephalon. When disease has made the expression of emotion impossible by destroying the motor innervation from the diencephalon, there may remain, as we have already seen, the possibility of emotional feeling. That feeling must be central in origin, and on the proposed theory it is occasioned by afferent impulses from an intact thalamus. But when the diencephalon is widely involved in a morbid process, as in the cases of Fulton and Bailey, both the expression and the feeling disappear together.

The greatest service this new theory can do is to divert the experimental study of the emotions from the periphery to the brain.

REFERENCES

1. Alvarez, W. C. Ways in which emotion can affect the digestive tract. *J. Amer. Med. Asso.*, 1929, **92**, 1231-1237.
2. Bainbridge, F. A. The physiology of muscular exercise. (2nd ed., rev. by G. Anrep) London: Longmans, Green, 1923. Pp. viii+226.
3. Bard, P. A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *Amer. J. Physiol.*, 1928, **84**, 490-515.
4. Bazett, H. C., & Penfield, W. G. A study of the Sherrington decerebrate animal in the chronic as well as the acute condition. *Brain*, 1922, **45**, 185-265.
5. Bekhterev, V. Die Bedeutung der Sehügel auf Grund von experimentellen und pathologischen Daten. *Virchow's Arch. f. pathol. Anat.*, 1887, **110**, 322-365.
6. Bickel, A., & Sasaki, K. Experimentelle Untersuchungen über den Einfluss von Affekten auf die Magensaftsekretion. *Dtsch. med. Woch.*, 1905, **31**, 1829.
7. Britton, S. W. The prepotency of medulliadrenal influence in emotional hyperglycemia. *Amer. J. Physiol.*, 1928, **86**, 340-352.
8. Bulatao, E., & Cannon, W. B. The rôle of the adrenal medulla in pseudaffective hyperglycemia. *Amer. J. Physiol.*, 1925, **72**, 295-313.
9. Campos, F. A. de M., Cannon, W. B., Lundin, H., & Walker, T. T. Some conditions affecting the capacity for prolonged muscular work. *Amer. J. Physiol.*, 1929, **87**, 680-701.

10. Cannon, W. B. The mechanical factors of digestion. London: Arnold, 1911. Pp. xi+227.

11. ————. Bodily changes in pain, hunger, fear and rage. New York: Appleton, 1915, 1920. Pp. viii+311.

12. ————. The isolated heart as an indicator of adrenal secretion induced by pain, asphyxia and excitement. *Amer. J. Physiol.*, 1919, **50**, 399-432.

13. ————. The James-Lange theory of emotions: a critical examination and an alternative theory. *Amer. J. Psychol.*, 1927, **39**, 106-124.

14. ————. Die Notfallsfunktionen des sympathico-adrenalen Systems. *Ergeb. d. Physiol.*, 1928, **27**, 380-406.

15. ————. Neural organization for emotional expression. Feelings and emotions: the Wittenberg Symposium. Worcester, Mass. Clark Univ. Press, 1928. Pp. 257-269.

16. Cannon, W. B., & Britton, S. W. Pseudaffective medulliadrenal secretion. *Amer. J. Physiol.*, 1925, **72**, 283-294.

17. ————. The influence of motion and emotion on medulliadrenal secretion. *Amer. J. Physiol.*, 1927, **79**, 433-465.

18. Cannon, W. B., Lewis, J. T., & Britton, S. W. A lasting preparation of the denervated heart for detecting internal secretion, with evidence for accessory accelerator fibers from the thoracic sympathetic chain. *Amer. J. Physiol.*, 1926, **77**, 326-352.

19. ————.. The dispensability of the sympathetic division of the autonomic nervous system. *Boston Med. & Surg. J.*, 1927, **197**, 514-515.

20. Cannon, W. B., McIver, M. A., & Bliss, S. W. A sympathetic and adrenal mechanism for mobilizing sugar in hypoglycemia. *Amer. J. Physiol.*, 1924, **69**, 46-66.

21. Cannon, W. B., Querido, A., Britton, S. W., & Bright, E. M. The rôle of adrenal secretion in the chemical control of body temperature. *Amer. J. Physiol.*, 1927, **79**, 466-507.

22. Cannon, W. B., & Rapport, D. Further observations on the denervated heart in relation to adrenal secretion. *Amer. J. Physiol.*, 1921, **58**, 308-337.

23. ————. The reflex center for adrenal secretion and its response to excitatory and inhibitory influences. *Amer. J. Physiol.*, 1921, **58**, 338-352.

24. Cushny, A. R. Pharmacology and therapeutics. (9th ed., rev. by C. W. Edmunds & J. A. Gunn). Philadelphia: Lea & Febiger, 1928. Pp. x+743.

25. Dana, C. L. The anatomic seat of the emotions: a discussion of the James-Lange theory. *Arch. Neur. & Psychiat.*, 1921, **6**, 1-8.

26. Dumas, G. Le choc émotionnel (réactions glandulaires et musculaires). *J. de psychol.*, 1928, **25**, 130-164.

27. Dusser de Barenne, J. G. Récherches expérimentales sur les fonctions du système nerveux central, faites en particulier sur deux chats dont le neopallium a été enlevé. *Arch. néerl. de physiol.*, 1920, **4**, 31-123.

28. Elliott, T. R. The control of the suprarenal glands by the splanchnic nerves. *J. Physiol.*, 1912, **44**, 374-409.

29. ————. The innervation of the adrenal glands. *J. Physiol.*, 1913, **46**, 285-290.

30. Fulton, J. F. Muscular contraction and the reflex control of movement. Baltimore: Williams & Wilkins, 1926. Pp. xv+644.

31. Fulton, J. F., & Bailey, P. Tumors in the region of the third ventricle: their diagnosis and relation to pathological sleep. *J. Nerv. & Ment. Dis.*, 1929, **69**, 1-25, 145-164, 261-277.

32. Gaskell, W. H. The involuntary nervous system. London: Longmans, Green, 1920. Pp. viii+178.

33. Goltz, F. Der Hund ohne Grosshirn. *Pflüg. Arch. f. d. ges. Physiol.*, 1892, **51**, 570-614.

34. Head, H. Studies in neurology. Vol. 2. London: Oxford Univ. Press, 1920. Pp. viii+333-862.

35. ———. Release of function in the nervous system. *Proc. Roy. Soc.* London, 1921, **92**, 184-209.

36. Houssay, B. A., & Molinelli, E. A. Centre adrénalino-sécréteur hypothalamique. *C. r. Soc. biol.*, 1925, **93**, 1454-1455.

37. Izquierdo, J. J., & Cannon, W. B. Emotional polycythemia in relation to sympathetic and medulliadrenal action on the spleen. *Amer. J. Physiol.*, 1928, **84**, 545-562.

38. James, W. The physical basis of the emotions. *Psychol. Rev.*, 1894, **1**, 516-529.

39. James, W., & Lange, C. G. The emotions. Baltimore: Williams & Wilkins, 1922. Pp. 135.

40. Karplus, J. P., & Kreidl, A. Gehirn und Sympathicus. *Pflüg. Arch. f. d. ges. Physiol.*, 1909, **129**, 138-144; 1910, **135**, 401-416; 1911, **143**, 109-127; 1918, **171**, 192-200; 1927, **215**, 667-670.

41. Kirilzev, S. [Cases of affection of the optic thalamus.] (Russian.) Reviewed in *Neur. Zentbl.*, 1891, **10**, 310.

42. Kodama, S. Effect of stimulation of the sensory nerves upon the rate of liberation of epinephrine from the suprarenal glands. *Tohoku J. Exper. Med.*, 1923, **4**, 166-242.

43. Langley, J. N. The sympathetic and other related systems of nerves. E. A. Schäfer's Text-book of Physiology. Vol. 2. London: Pentland, 1900. Pp. 616-696. See also: Das sympathische und verwandte nervöse Systeme der Wirbeltiere (autonomes nervöses System). *Ergeb. d. Physiol.*, 1903, **2**, 818-872.

44. ———. The autonomic nervous system. Part I. Cambridge: Heffer, 1921. Pp. viii+80.

45. Magnus, R. Körperstellung. Berlin: Springer, 1924. Pp. viii+740.

46. Pavlov, I. P. The work of the digestive glands. (2nd ed.) London: Griffin, 1910. Pp. xiv+266.

47. ———. Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927. Pp. xv+430.

48. Rothmann, H. Zusammenfassender Bericht über den Rothmannschen grosshirnlosen Hund nach klinischer und anatomischer Untersuchungen. *Zsch. f. d. ges. Neur. u. Psychiat.*, 1923, **87**, 247-313.

49. Satáḱé, Y., Watanabé, M., & Sugawara, T. Effect of fastening and sensory stimulation upon the rate of epinephrine output from the suprarenal gland in dogs. *Tohoku J. Exper. Med.*, 1927, **9**, 1-40.

50. Schafer, E. S. The endocrine organs. Part I. (2nd ed.) London: Longmans, Green, 1924. Pp. ix+175.

51. Schiff, M. Moritz Schiff's gesammelte Beiträge zur Physiologie. Vol. 3. Lausanne: Benda, 1896.

52. Sherrington, C. S. The integrative action of the nervous system. New Haven: Yale Univ. Press., 1906. Pp. xvi+411.

53. Soury, J. Le système nerveux central. Paris: Carré et Naud, 1899. Pp. x+1863.

54. Stewart, G. N. The significance of the suprarenal glands in relation to the vital processes. Endocrinology and metabolism. Vol 2. New York: Appleton, 1922. Pp. 127-168.

55. Sugawara, T., Watanabé, M., & Saito, S. Effect of stimulation of the sensory nerves on the rate of liberation of epinephrine from the suprarenal glands. *Tohoku J. Exper., Med.,* 1926, **7**, 1-79.

56. Wiggers, C. J. Circulation in health and disease. (2nd ed.) Philadelphia: Lea & Febiger, 1923. Pp. viii+662.

57. Wilson, S. A. K. Pathological laughing and crying. *J. Neur. & Psychopathol.,* 1924, **4**, 299-333.

58. ————. Modern problems in neurology. New York: Wood, 1929. Pp. 364.

59. Woodworth, R. S. & Sherrington, C. S. A pseudaffective reflex and its spinal path. *J. Physiol.,* 1904, **31**, 234-243.

CHAPTER 13

EMOTION: II. THE EXPRESSIONS OF EMOTION

CARNEY LANDIS
Wesleyan University

Three methods have been available for the experimental investigation of emotion: (1) the study of the expressions which accompany—or are—the emotion, (2) the method of descriptive and introspective reports of the experiential phenomena of "mental life" which are characterized by feeling, and (3) the investigation of the action of the nervous system and of the endocrine glands. Of these methods, one may say that the first has furnished the greatest mass of factual material, the second is the earliest historically, but has been the least productive, while the last is undoubtedly the most fundamental since it is from neuro-endocrine physiology that psychology must draw its ultimate explanations of emotion.

The major portion of this chapter will be given to the direct consideration and evaluation of our present knowledge concerning the expressions of emotion in the narrow sense of the term. The material which is available on the description of affective processes in mental life will be considered briefly under the heading of "Feeling." The neural basis of emotion has been taken up in the previous chapter and will be considered in this chapter only in so far as it is to be properly considered as part of the expressive mechanisms.

An attempt will be made to adhere to the following definitions of terms:

Emotion: The sum total of the experiences during any period in which marked bodily changes of "feeling," "choc," or "upset" take place.

Feeling: (*a*) Experience less vivid than emotion. (*b*) The report of vaguely perceived sensations most probably derived from sensory impulses over the autonomic nervous system.

Choc: The poorly coordinated responses elicited by stimuli not reacted to by the organism in its immediate past and which occur suddenly. The term will be used to denote the behavior usually called "surprise," "start," or "shock," but without the mental connotations of these words.

Upset: The condition of an organism brought about by a series of stimuli which cannot be reacted to adequately, with consequent dissociation of neural function and of various usually integrated behavior patterns.

Expression: Any change in an organism which may be said to be a criterion of, accompaniment of, or part of a reaction system. This change must be noted by a second organism or by some recording or measuring instrument before it may be truly considered as an expression of emotion.

Affect or Affective Experience: A blanket term used to denote any variety of emotional experience or emotional concomitant.

Criteria

How do we recognize an emotion when we see it? What are the marks which characterize one expression as "emotional" and another as "language"? Is such recognition innate or acquired? What are the valid criteria of emotion or of the expressions of emotion? Can any reaction which is accompanied by pleasantness or unpleasantness be classified as emotional or as an affective experience? Is emotion an entity of itself, either in the sense of a pattern or series of patterns, of physiological functions of a specific variety or in the sense that it is a unique experience of mental life? Anyone who gives the matter some thought can easily add a dozen or more such questions which should have some sort of an answer in the systematization of our knowledge on this subject. Many such questions have been answered by experiments. Others depend at present upon rational deductions for an answer. Still others have not as yet found a satisfactory answer.

The behavior of the human infant indicates that the recognition of emotional expression is not present at birth but develops very rapidly, as if there were some innate tendency awaiting only maturation and directional stimuli. The development of language habits from the random verbalizations of the infant is probably very closely analogous to the development of the ability to recognize and to use expressions of emotion. This seems to be less and less true as one goes down the animal scale. Lower animal forms do not react differentially to expressions but only to the direct application of stimuli of a nocuous or benign character.

Socially, the recognition of expressions of emotion is clearly of an acquired nature. This acquisition is of the nature of a language mechanism. That is, the recognition is learned just as verbal speech is learned. Some individuals become quite adept at one or either; others do not. The expressions themselves, so far as the facial expression is concerned, are *usually* learned responses. Just as one learns to say "Oh" or "Ouch" in response to injury or surprise, so the expression accompanying the verbalization is *usually* a learned response. The entire question of distinction of acquired expression from native or innate expression of emotion is one which is not clearly defined, but which will be considered at greater length later in this chapter.

The expressions of emotion are diverse and include any portion of the body which is capable of reacting or having part in a reaction. Some of the more common reactions which have been considered expressions of emotion are as follows: facial expression, blushing or flushing, certain patterns of bodily movement, certain types of verbal responses, laughing and crying, changes in the secretion of sweat, blood-pressure, blood volume, respiration, changes in the activity of the gastro-intestinal system, variations in metabolism, changes in the balance between the secretions of the endocrine glands, variations in balance between portions of the nervous system, electrical responses, and so forth.

Facial Expression and Vocalization

Historical.[1] The earliest systematic study of facial expression is that of physiognomy. Physiognomy in its original scientific sense was a twofold science: (1) a mode of discriminating or predicting character by the outward appearance, and (2) a means of classifying form and feature. The first of these uses is, of course, outside of the scope of modern science. The classification of the relationship between facial expression and character is found in the earliest literature of the Greeks and of the Romans. The first systematic treatise is that of Aristotle, who devotes six chapters to the consideration of the method, the general signs of character, the particular appearances characteristic of the dispositions, of strength and weakness, of genius and stupidity, etc. The practice of physiognomy was one of the first branches of intellectual interest through all of the Dark and Middle Ages. Numerous treatises showing shrewd observation mixed with mysticism and pious teachings are to be found in the extant literature of those days. This work is best characterized by the physiognomical publications of Lavater, a Swiss minister. He attempted to mix piety with character observation and shrewd deductions as to the relationship existing between physique (physiognomy) and character or personality. His observations, though interesting, are of little scientific importance.

After the many centuries of speculation concerning the importance of facial expression as a criterion of emotion or character, Darwin was able to furnish a new hypothesis and systemization on which it has been possible to carry out experimental work. In 1873, Darwin (11) announced three principles which he hoped would account for most, if not all, of the expressive gestures involuntarily used by man and the lower animals in emotion. These principles are as follows:

1) *The Principle of the Serviceable Associated Habits.* Complex movements which may under certain circumstances be directly or indirectly useful are retained after the use has passed. The peculiar "pawing" movements of the cat's forefeet when pleased are relics of the purposeful use of the same movements to start or increase the flow of milk from the mammary glands of the mother. Such movements were associated with a pleasurable and satisfied feeling and tend to recur when such a feeling recurs. Wundt considers that this is but a special case of Darwin's third principle (*q.v.*).

2) *The Principle of Antithesis.* Darwin states, "Every movement which we have voluntarily performed throughout our lives has required the action of certain muscles; and when we have performed a directly opposite movement, an opposite set of muscles has been habitually brought into play . . . . So when actions of one kind have become firmly associated with any sensation or emotion, it appears natural that the actions of a directly opposite kind, though of no use, should be unconsciously performed through habit and association under the influence of a directly opposite sensation or emotion." Impotence is expressed by raised eyebrows, shrugged

---

[1] In part modified from Titchener (46).

shoulders, and open palms, these being antithetical to the frowning brow, thrown-back shoulders, and the clenched fists, symbolic of rage and power. James has commented, "No doubt a certain number of movements can be formulated under this law; but whether it expresses a causal principle is more doubtful. It has been by most critics considered the least successful of Darwin's speculations on this subject."

3) *The Principle of the Direct Action of the Nervous System.* The sudden release of large quantities of nervous energy in choc demands unusual outlets from the central nervous system. This overflow takes place according to the innate connections within the brain. The results are shown in a general disturbance of organic functions due to the exciting or inhibitory effects of this irradiation of energy. By this principle are explained such phenomena as the muscular tremors in fear or great joy, the increased glandular activity of the liver and kidneys, the alterations in cardiac and vasomotor functions, and such movements as the clapping of the hands in joy. This principle is admitted by James and is practically identical with Wundt's first principle.

Wundt like Darwin has three principles of systematization.

1) *The Principle of Direct Change in Innervation.* This is but a different wording of Darwin's third principle, involving the hereditary transmission over certain nervous connections, as the reflex of weeping. The transmission of a characteristic family physiognomy or general expressive attitude is very common; and we invariably argue from the physical similarity to a similarity of mood, even though there may be no direct possibility of imitation of the parents by the children. The frightened man stutters, not merely because his tongue mechanically refuses service but also because his thoughts are really brought to a standstill, showing the effect of reaction on the ideational processes in emotion.

2) *The Principle of the Association of Analogous Sensations* of similar feeling-tone which mutually reinforce. This process forms the basis of the most characteristic of all emotional expressions, the "mimetic movements." These are physiologically conditioned by reflex movements in and about the facial sense-organs; thus the expression which stands for "bitter" is an arrangement of the parts of the buccal cavity most sensitive to bitter in such a way as to prevent their excessive stimulation by the unpleasant taste. The "sweet" expression, on the contrary, is that calculated to favor the continued stimulation of the tip of the tongue, the part most sensitive to sweet substances. By virtue of the second principle, these reactions have come to appear in response not only to an actually bitter taste but also to an emotional condition which possesses the same general feeling-tone. So the expression has become symbolic. A wry face may denote a mental as well as a physical pain.

3) *The Principle of the Connection of Movement with Sense Ideas* explains all the facial expressions and gestures which are not included under the two previous principles. Here belong movements of the arms and hands, the clenching of the fists in anger, and also certain facial expressions such as the curling of the lip in scorn, and the staring eyes of

surprise. This principle is closely allied to Darwin's first. It is not to be thought that any complex expression of emotion must be explained by a single principle. Such phenomena as laughing and weeping demand the use of all three principles for their elucidation.

James (22) proposes the five following principles: (1) The weakened repetition of movements which formerly were of utility to the subject. This is equivalent to Darwin's first proposition. (2) The principle of reacting similarly to analogous feeling stimuli. This is identical with Wundt's second principle. (3) The principle of weakened repetition of movements which under other conditions were physiologically necessary effects. The respiratory disturbances of anger and fear, e.g., may be considered as "organic reminiscences of the blowings of the man making a series of combative efforts, of the pantings of one in precipitate flight." (4) The principle of the mechanically determined idiopathic effects of the stimulus, i.e., the physiological outpourings of excess of nervous energy through the easiest drainage channels (cf., Wundt's first and Darwin's third law). (5) The principle of the mechanical perpetuation of emotional reactions which may be called accidental as far as their origin goes. For some of our emotional reactions no plausible reason can be conceived. "In fact, in an organism as complex as the nervous system there must be many such reactions, incidental to others evolved for utility's sake, which would never themselves have been evolved independently, for any utility they might possess." In conclusion it may be said that no one of these explanatory series of principles is logically complete. There is need of further observation, and perhaps of a new construction of principles upon the basis of the psychology of action.

It might be well to note here that there is a very pertinent criticism of Darwin's, Wundt's, or James' descriptions of expressions of emotion which are based on folk psychology or the study of the various races of mankind. The reports on which these principles are based were made for the most part by individuals trained in the habits of our western culture, that is to say, their powers of observation and methods of perception as well as of interpretation were those of western European culture and hence the expressions reported would be of that nature.

Anatomy and Physiology. In 1806 Sir Charles Bell published the first edition of his *Anatomy and Philosophy of Expression.* This treatise lays the foundation for the scientific investigation of facial expression. Bell developed the notion that the expressive reactions of the face were based upon the anatomical relationships of the muscles so that expression was limited in its scope by the anatomical structure. In a way his ideas of the nature of emotion and the expressions of emotion are forerunners of Darwin and of James in that he insists on the bodily expression as a basic fact of emotion and takes up in detail the changes in respiration, circulation, and the like which have seen so much recent experimental investigation. Duchenne in 1862 published *La Mécanisme de la Physiognomie Humaine.* This work is basic for the physiological study of facial expression. Duchenne's method was that of stimulating each muscle or muscle group of the face directly by means of a galvanic current so obtaining the expressions, many of which had been viewed as complicated responses. The resulting reactions were photographed, and from the examination of the pictures the thesis was developed that expression depends primarily upon the activity of single muscles.

Frappa (17) has given an opposing view to that of Duchenne which has been very interestingly developed by Dumas (13). Frappa's contention is that facial expression may be reduced to three fundamental expressions: astonishment, grief, and joy. From these three expressions he believed that one can synthetically build up most, if not all, other human expressions. Dumas has reworked the hypothesis of Frappa and shown in a most suggestive fashion that the three elementary expressions are easily explained in terms of direct neural function. The expression of surprise (astonishment) Dumas considers to be due to the inhibition in choc of the innervation of the muscles and particularly the innervation of the masseters and the orbicular muscles of the eyes—which is sufficient to account for the expression termed "surprise." The expression of joy is to be explained by the raising of the tonicity of the entire body including that of the face. Hence joy may be said to be the pattern of heightened facial tonus. Grief is to be considered as the opposite of the tension of joy; that is, the facial muscles are relaxed and the balance of forces between the stronger (heavier) muscles as compared to the weaker muscles results in the expression called grief.

The act of blushing or flushing is a distinctive facial reaction which enters into the physiology of expression. The fact that we use two terms, blush and flush, to describe the same actual facial phenomenon indicates that the reaction itself is not usually separated from the total configuration. Blushing may be said to be a peculiar human trait since no animals show this type of reaction. The physiology of blushing will be discussed in more detail under the head of circulatory expressions. It should be pointed out, however, that it is an element in facial expression which is separate and distinct from the muscular activity usually considered the sole functional element of facial expression. The mechanism of this circulatory disturbance is quite distinct from the mechanisms which control the muscular reactions. The fact that there is an association between these responses indicates something of the complexity of the problem with which we are dealing.

Interpretation of Facial Expression. Both historically and individually the human race has a great trust in its ability to read and interpret facial expression. The historical treatment of expression is that of physiognomy; the biological treatment, that which grew out of the theory of evolution, particularly the Darwinian attempt to demonstrate the evolution of mental traits. It must be distinctly understood that such reactions are *social expressions* and are of the same nature as expression in language or in any other acquired activity.

One may find not only in the older literature but in very recent texts of psychology elaborate systematic classifications of expressions or of emotions. (Usually the distinction between expression and emotion is a vague one at the best.) These classifications or dictionaries are of value in calling attention to the diversity of reactions of the face. Indeed they indicate that the face is second only to the vocal mechanism in its ability to give finely differentiated reactions. It remains to be seen whether or not such an assumption is based upon fact.

The acquisition of the ability to interpret facial expression has been studied by Gates (18). The method used was the presenting to children of different age levels a series of posed photographs depicting the traditional expressions of emotion, while asking the child to name the expression on these photographs. The study indicated that the success which a child may have in acceptably naming expressions is closely correlated with his age level, but, even at best, this interpretive process is far from perfect. The experiment also shows that this perception of differences in expression is clearly a social acquisition of the individual.

The work upon the interpretation of photographs which had been posed as expressions of emotions was initiated by Miss Feleky (15). She herself posed for a series of pictures which she hoped would run the entire range of expression and which various judges would agree were typical expressions of this or that emotion or emotional situation. She found that the expressions of disgust, sneering, and breathless interest were judged correctly more often than any others, while those of suspicion, religion, anger, fear, hate, and rage were judged very poorly. Several years later Langfeld (33) tried a similar experiment making use of an artist's sketches of posed emotional expressions. He found that laughter could usually be interpreted correctly, that fear and anger were confused, that surprise and suspicion were confused, and that the contempt-scorn group was judged more correctly than any other expression. His subjects reported that the use of kinaesthetic imitation and imagination of the situation aided in the labeling of the photographs. A check of the reliability of this experiment is given in a second study in which positive and negative suggestions were given to the judges. The factor of individual suggestibility of different subjects colors the results, but in general it is shown that variations in positive suggestion are a distinct aid, while negative suggestion interferes with all judgments so markedly as to indicate a clear sense of distrust of one's own judgment on the part of the judge. Ruckmick has carried out a further study with a third series of photographs, confirming for the most part the results of Feleky and of Langfeld. Buzby (7), using an adjustable "Piderit" model or caricature of a face, found that only two per cent of the judgments of a group of 716 subjects agreed with the title of anger which had been applied to one particular combination, while 61 per cent agreed with the title of horror as applied to another combination. These two expressions represent the extremes on his scale of agreement.

These studies all smack of guessing games. The game is that of attaching a label to a photograph so that the label will agree with the notion which the experimenter held concerning the picture. They are in no sense true studies of the expressions of emotion but are investigations of the socialization of the perception of facial contortions. Their value lies in the fact that they show clearly that even the most conventionalized of social expressions is very poorly interpreted by the average observer. So poor are the judgments, that little trust or reliability may be placed upon the judgments when the observer has *only the face* to guide him. Note the phrase "only the face," for it is practically certain that if the observer could see the entire situation his judgment would be much more accurate.

The study of emotional expressions in real situations of everyday life has never been satisfactorily carried out under controlled conditions. Mantegazza has published vivid descriptions of the behavior of criminals on the way to execution, maniacal patients, and other individuals under the influence of intense emotion. His descriptions were thoroughly qualitative and follow the traditional literary lines of description. It is theoretically possible to make a study of the expressive reactions of a group who have been cinematographed during some situation of high affective value, but the research still waits the doing.

McKenzie has made some very interesting studies of facial expression of athletes during strenuous competition and after partial collapse from such competition. He bases his work on photographs which were reproduced in sculpture. The expressions are marked by extreme tonicity of the face in strenuous situations, and, in collapse, by the blankness of lowered muscular tonus.

Experimental Studies of Facial Expression. Landis has conducted several experimental studies of facial expression. In the first of these he was chiefly interested in the recording and analyzing of the expressions of persons subjected to a controlled series of situations of a more or less emotional nature. The expressions were recorded both photographically and cinematographically. The first of the studies showed there was a great deal of variability in the expressions obtained. So long as the expressions were those which accompanied a mild degree of stimulation they closely approximated the reactions traditionally considered appropriate to the situation. When several of the subjects who had been very expressive in the original situations were asked to return and to endeavor to reproduce the expression which they believed that they had given in the original situations, it was found that, in general, these attempts resulted in the purely traditional expressions rather than those which were actually obtained in the more severe of the emotional situations.

A second study by Landis repeated in part the first experiment. The situations were made more intense and real, so that there was little doubt that true emotion was elicited. Facial expression was recorded by means of a camera capable of following fairly rapid facial reactions. Records were obtained of blood-pressure and of the respiratory rates during the entire series of situations. The stimulation evoked emotional disturbance in practically every subject, while certain subjects showed signs of complete upset. Twelve men and 12 women, advanced undergraduate or graduate students served as subjects, all undergoing the same series of events. In all, about 1200 photographs were taken.

These photographs were analyzed by a semi-objective determination of the amount of movement shown by the principal groups of facial muscles. The emotional expressiveness was determined by an approximate measurement of the amount of movement of facial muscles, and in terms of the subjective significance which the facial pattern conveyed to the experimenter. It was found that all individuals tend to use certain muscles or muscle groups (expressions) in the majority of their expressive reactions and to exclude other muscle groups. There are marked individual variations in the expressive reactions so produced. In no situation was a contraction of a muscle, group of muscles, or expression found to occur with sufficient frequency to be significant. That is, there was no expression typical of any situation in this experiment. Nor was any expression typically associated with any verbal report given in this experiment. Some degree of smiling was the most common facial reaction, occuring in 34 per cent of the photographs. Asymmetrical expressions almost never occurred as emotional expressions. The men in this experiment used more facial reactions than the women.

When certain of the subjects were recalled and asked to endeavor to reproduce the feelings or emotions which they experienced (centrally aroused emotion) in the original situation, it was found that no expression or pattern of response characterized even these imagined situations or experiences.

Ranking the amount of facial movement which accompanied the various emotions named in the verbal reports, revealed that pain, surprise, anger, exasperation, crying, disgust, sex, and revolting gave varying amounts of expressive movements, in the order named. That each subject has two or three facial patterns which he uses in all situations and which constitute the vast majority of that individual's expressiveness is an interesting point bearing on a general theory of emotion. In general, simple emotions or simple expressions were never found; they were always complex.

The emotional responses of infants has been studied by Sherman (43). He used four types of stimulation: hunger, dropping, restraint, and pain. When motion pictures of these reactions were shown to students, he found that 12 to 25 different emotions were named as appropriate titles for the pictures. When nurses and medical students were shown three types of reactions resulting from dropping, restraint, or pain, without seeing the actual stimulation, the nurses named seven emotions and medical students eight. The difference in number of titles applied is probably a function of the size of the individual's vocabulary. If the pictures were accompanied by a knowledge of the actual stimulation, the resultant reaction was judged with a great deal more unanimity than under any other conditions. The basis of judgment stated by the observers indicates, for the most part, that the differentiation was strongly influenced by the knowledge of the stimulus. This was further shown by an additional experiment in which the stimuli were attached in the motion picture to reactions elicited by other stimuli. The interpretation of these pictures indicated that the stimulus preceding the reaction shown was usually the deciding factor in the name given to the reaction.

Sherman's work indicates clearly that no reliance can be placed upon the judgments of either trained or untrained observers of the expressive reactions of the infant. If, as Watson and others have stated, anger, fear, and love are the three innate patterns of emotional response, then Sherman should have found a large percentage of agreement among the judges in the naming of reactions. Whether the reaction was named rage, fear, or love is of little import, but the judgments should fall into three clearly marked groups. This was not found. Sherman's work shows that the differentiations of the responses of the infant are differentiations made when the stimulating conditions are known, and are accurate only when these conditions are known.

The experimental investigations of the facial expressions of emotion of the adult or the bodily activity of the infant indicate that there is no pattern of expression which may be said to characterize any situation or emotion of any one individual or group of individuals. Practically all of the current theories of emotion demand that we should find such patterns, if not in the voluntary musculature, then in the visceral responses of the individual.

Vocal Expression. The significance of variations in the vocal expression as concomitant to, or part of, emotion is older phylogenetically than is facial expression. Darwin classifies most of the vocalization of animals as emotional. The practical significance of variations in speech conveying emotional meaning which may or may not agree with the words is common knowledge. There is, in view of these facts, surprisingly little experimental work available which classifies these data or explains the phenomena. Sherman's experiment (43) in which a group of observers were presented with the *cries* of infants stimulated by hunger, dropping, restraint, or pricking with a needle is of interest in this connection. These observers named 12 different emotions which showed very little relationship between the name given and the qualitative nature of the stimuli. When a group of students heard only the cries of the infants, about 20 per cent of the judgments were colic. On the other hand, when a similar group of observers were presented with motion pictures of the reactions in similar stimulating conditions, not one judgment of colic was made. Evidently the cries were associated by the observers with the organic processes of the infant rather than the nature of the emotions. The intensity and the duration of the cries were the determining factors in judgments of the observers. The more intense the stimulating condition the sharper was the ensuing cry and the longer was its duration.

Gates (19) has studied the development of the ability in school children of various age groups to interpret correctly the auditory element of expression. Phonographic records were made of the recitation of the alphabet in tones of happiness, unhappiness, anger, fear, surprise, etc. When this record was given before 627 children, the following results were obtained: "(1) an increase in the capacity for understanding with age, grade, and school experience; (2) the superiority, in this test, of the more intelligent over the less intelligent school children; and (3) differences in the ease of interpretation of the types of auditory stimuli."

Seashore (42), by means of phonophotography, has analyzed the peculiar attributes of music or speech which are regarded as expressive of emotion and finds that the presence of vibrato (defined as a synchronous oscillation of pitch and of intensity at a rate of from five to eight per second) in artistic singing and in speech gives the element which we commonly term emotional.

Laughter and Crying. Two varieties of expressive facial reactions are certainly innate. No one has ventured to criticize the notion that laughter or crying were pattern responses. The criticism has been successfully urged that very frequently these responses are not expressions of emotion. The treatment which Dumas (13) has given to the problems of laughter and of crying is most illuminating and will be followed in part in the discussion of this subject.

There are, according to Dumas, five major problems in connection with laughter: the problem of its physiological mechanism, the problem of *le rire de la joie,"* the problem of *"le rire du comique,"* the problem of the connection between the mental state and its outward expression, and the problem of laughter as a social phenomenon. The physiological mechanism for laughter consists of the 15 facial muscles which are used in smiling, the respiratory muscles, the vocal mechanisms, and, in extreme cases, other muscles throughout the body.

There is a fundamental difference between the joyful laugh and the comic laugh. The joyful laugh is a sort of bubbling over of good humor. It is very common in children. It is usually seen when some restraint is suddenly removed, as when children come out of school, or when there is some sudden good fortune, as when old friends meet after a long separation. The laughter caused by tickling is probably to be classed along with these; also the laughter caused by "laughing gas," alcohol, or nervous disorders.

The comic laugh is always a laugh *at* some joke or ludicrous situation, and it may represent an entirely different attitude from that which produces the joyful laugh. The nature and cause of this attitude has been a problem for philosophers ever since Aristotle. The question, "Why are things funny?" is as difficult as the question, "Why are things beautiful?" According to Aristotle, laughter is due to "the sudden feeling of triumph which comes with the sudden perception of a superiority in us, by comparison with the inferiority of others or our own former inferiority." When another person slips on a banana peel, or sits down on a non-existent chair, or commits a social blunder, we laugh because of the instinctive feeling that "*I* didn't do that!" Hobbes restated the same general idea when he said that laughter is due to the "degradation of some person or interest possessing a certain dignity." Examples of this are the dog in church, the wedding procession spattered with mud, and the fit of sneezing in the middle of a proposal of marriage. For Kant, laughter was due to the sudden relief of tension or expectation. We laugh because of relief of tension when a clown runs towards a horse as if to leap in the saddle at one bound but stops short to flick a microscopic speck of dust off the bridle. For Schopenhauer, humor consisted in the sudden perception of a discord between reality and our idea or representation of reality. We laugh at a monkey or a cartoon because it is such a pitiful imitation of a man. According to Bergson, humor consists in the fact that, in place of an intelligent and well-adapted reaction, some individual makes an unintelligent and poorly adapted reaction to a situation. Humor is an infinite variation on the theme *"du mécanique plaque sur le vivant."* The customs officer who asked the shipwrecked sailor if he had anything to declare, the short man who stooped to go through the high doorway, the man who slipped on a banana peel, are all examples of this.

The fourth problem connected with laughter is that of finding the connection between the mental state (sum total of implicit and explicit responses ?) and outward expression. Given the feeling that a certain thing *is* funny, why should we express that feeling by working the muscles of the face? This is largely an unsolved problem, and necessarily depends on the actual nature of the mental state. Spencer explains laughter as an outlet of surplus energy. This agrees with the theories of Hobbes and Kant, but not very well with the others. Dumas suggests that when there is a sudden contrast or contradiction, especially when the contrast is a descent from the greater to the less, the nervous energy stored up in the first case finds an unexpected outlet through laughter.

With regard to the social aspect of laughter, Dumas calls attention to its contagious nature. We laugh far more easily in a group than when alone. Laughter, like facial expression, has become a kind of language. Sometimes we may laugh merely to show others that we understand the joke. At other times it is a communication of real good will and the spirit of fun, usually, perhaps, both. We sometimes laugh for the pure joy of laughing. For instance, in certain situations the slightest stimuli will start us off and prolong the laughter out of all reason, as the contagious giggling of the school child.

The question of tears involves four problems: physiological, psychological, psycho-physiological, and social. The physiological question is, as in the case of laughter, comparatively simple. The lachrymal glands, situated over the outer corners of the eyes, send their lubricating fluid over the eyeballs through four or five large and eight to ten small canals. These glands function continuously when the eyes are open, since the action of the air on the eyeballs stimulates them to reflex action.

Tears (abnormal secretion of the lachrymal fluid) are produced in physical pain, grief, anger, extreme joy, and excess of laughter. The common element in the psychological situation seems to be physical stress. Tears, in the infant, probably occur as a sort of by-product of the general bodily excitation: the acceleration of the heart, the heightening of blood-pressure, the dilation of blood-vessels near the skin, and acceleration of breathing. The nerve-centers which govern tears are located in the opto-striate region, which also governs laughter. The anatomical approximation probably explains the physiological coincidence.

Tears, like facial expression and laughter, are used as language. Infants very early learn that crying often gets them what they want. Though the process is difficult to manage voluntarily, because it is originally reflex, it can be done. It is interesting to speculate whether other reflex glandular activities, such as the production of saliva or gastric juice, might have become language mechanisms if only they had been visible from the outside.

Borgquist (4) made a study of grief and crying by means of a questionnaire, scattered physiological data, and references from literature. He concluded that crying in the last analysis was a reaction which comes at the end of a period of stress where there has been strong effort and exhausted nervous energy. It is essentially a breaking down of adaptive mechanisms which are no longer adequate to meet environmental conditions. These are clearly two groups of symptoms which accompany different moments of the crying act. The first are those active movements of "calling," as represented by the vocalization of the infant or child. The second group includes the facial expression, sobbing, tears, etc., which are to be interpreted as withdrawing movements going back to a primitive rejection of food.

Freud has advanced a theory of wit and laughter which explains some of the difficulties of the connection between the muscular activity of laughter and the mental content of comedy. He states the theory as follows: "A fore-conscious thought is left for a moment in the unconscious elaboration and the results are grasped by the conscious perception. 'Wit' seeks to draw a small amount of pressure from the free and unencumbered activities of our psychological apparatus and later to seize this pressure as incidental gain." The pressure of wit originates in an economy of expenditure of inhibition, the comic from an economy of expenditure in thought, and humor from an economy of expenditure of feeling. Freud evidently considers the expressions themselves as native reaction patterns since he fails to raise the question. His theory accounts only for the mental content.

The Relation of Facial Expression to Emotion. The investigations together with the rational deductions which have been considered thus far lead one to the conclusion that facial expression is but a more or less variable part of emotional experience. The presence or absence of pattern in expression from the voluntary musculature of the face adds to or detracts from the entire emotional experience but is only one element in a group of forces and not a very important element at that. The important elements in emotional expression are the situation and the level of activity of the body. The three patterns of facial expression, surprise, joy, and depression, result from the condition of the facial musculature as direct responses to the stress of the situation. The particular character of the expression is, to the best of our present knowledge, dependent on the survival of primitive connections within the nervous system. It is probable that early in animal evolution the baring of the teeth and the opening of the mouth in periods of emergency was of survival value and that this particular connection has persisted. Still another probable explanation lies in the fact that the nucleus of the fifth or facial nerve is located in the mid-brain in close approximation to the centers of organization of affective response. This proximity might conceivably lead to the use of the face as an overflow mechanism in the draining off of nervous energy expended in affective experience.

Bodily Activity

The question of the activity of the entire body in affective experience is an exceedingly complicated one. The analysis of behavior in adaptive activity of the organism to an emotional situation shows that it is so highly complicated that any attempt to make a simple or orderly behavior classification can meet with but little success. Carr has pointed out that the bodily responses in emotion should be considered under three heads: (1) the *act,* which refers to those processes which constitute adaptation to the objective situation, (2) the *emotion,* which refers to those activities which increase effectiveness of the act, (3) the remaining processes which constitute the *by-products* of the emotion or the act. He goes on to show that there is no coordinate fourth group of processes which may be termed expressive activities, expressive responses of emotion for the body as a whole.

Even though there are no well-marked patterns of bodily responses in adult humans, it seems plausible that such responses might be found in animals or in infants. Darwin's studies, which have been mentioned above, seem to imply that there are well-marked emotional responses in animals. Subsequent work of the physiologists, notably Cannon (8) and Sherrington, shows that these total responses are almost as variable in the animal as in the adult human being. Emotional responses are limited only by the mechanical limits of the activity of the animal.

Several years ago Watson (49, pp. 108-158) reported the results of his studies concerning the emotional responses in the new-born infant. He stated that there were three well-defined patterns: rage, fear, and love. The names which he gives to these patterns are his own labels. (N. B., Sherman's work quoted above). The patterns of response which Watson described, he believed to be innate and adequate, in the evolutionary sense, to the situation. It is probable that it would be equally easy to define and limit three other patterns of response which might be regularly elicited by appropriate stimuli and which would have as great inherent worth as those to which Watson has called attention. The main point is that the appearance of true patterns of expression in the human adult or infant or in higher animal forms is a matter of extreme doubt. The response obtained by emotional stimulation is the determining factor.

The work of Freud and the psychoanalysts somehow suggests that in abnormal behavior there may be more or less well-defined bodily patterns of emotional expression. These, according to Freud, are for the most part symbolic or mimetic of repressed desires or wishes. The responses represent the acting out of "un-live-out-able" wishes. In the inescapable kinship of the human race the sameness in activity of all members of the race will appear in these mimetic or symbolic pattern responses, which are most clearly defined in neurotic individuals. These findings are all based on clinical observation and until they have been subjected to controlled experimental verification their significance in experimental psychology is questionable.

The most recent development in psychology, namely the *Gestalttheorie,* carries with it certain implications concerning emotional expression. This theory holds that emotional expressions are perceived as such, as totalities, and not in terms of the amount of muscular activity of the mouth, the wrinkling around the eyes, or in any other element analyzed out of facial response. Unpublished experiments which have been conducted by Landis and White show no constant tendency for either children or naïve adults to report the perception of these totalities. Rather the reports are given in terms of the amount of movement of some portion of the face. In such experiments it is, of course, clear that *set* or attitude is the all-important factor. It would be easily possible to set up an experiment in which Gestalt reports would be obtained. When the experiment is not planned so definitely, but the matter of perceptual report left open to the unguided individual, Gestalt reports are not usually obtained.

VISCERAL RESPONSES

Circulation, Blood, and Heart

1) *Blood volume.*[2] Following the original theoretical suggestions of Lange that the core of emotion was to be found in changes of blood distribution throughout the body, the changes in blood volume in affective experience have been extensively studied. Some modification of the Mosso plethysmograph, an instrument for determining changes in the volume flow of blood in any particular member or locality of the body, has been most frequently employed in this work. More than twenty experiments were reported between 1879 and 1925 which made use of the plethysmograph for analysis of the variations in blood volume in emotion. The plethysmograph does not lend itself to experimentation where the situation is intense or where the stimuli will call out marked voluntary bodily activity. Hence the blood-volume experiments have been chiefly directed towards the analysis of pleasantness, unpleasantness, or milder affective experiences in terms of blood changes. Most of these experiments note that there is an increase in the blood volume of the periphery of the body during pleasantness and a decrease during unpleasantness. However, certain experiments reporting contrary results to this general claim have been so well conducted as to leave the point in doubt. In general, there is little hope of the differentiation of pleasantness, unpleasantness, or of other affective experience on any basis of blood-volume changes in the body. Miss Eng (14) has recently reported an elaborate study of the emotional life of the child in terms of the plethysmographic records. Her work, in the main, confirms the indefiniteness of all of the previous find-

---

[2] See Eng (14) for technical diagrams and a discussion of technique.

ings. There was no more pattern in the emotional response of the child than of the adult. The "spontaneous" emotional reactions in children last longer and are stronger than those of the adult but are no more definite in their blood-volume pattern. Her work failed to confirm any of the standard psychological theories of feeling or emotion.

Weber (1910) carried on an experimental investigation of the changes in blood volume of the entire body. For this work he used the "balanced board of Mosso." He found that during pleasantness, sleep, or ideas of movement, there was a general flow of the blood to the periphery of the body. In all other situations there was a draining of the blood to the viscera and interior of the body. The chief criticism of the work of Weber is that, for the most part, his situations were artificial. He hypnotized and suggested to the subject the desired emotional situation. The question of the physiology and psychology of hypnotism and of the validity of expression or emotion during hypnotism is even more obscure than is the waking expression of emotion.

2) *Blood-pressure.*[3] The use of measurements of blood-pressure as an expression of emotion grew out of the work on blood volume. This work makes use of the sphygmomanometer, an instrument for the indirect determination of arterial blood-pressure. The measurements are usually made on the upper arm and expressed in terms of pressure in millimeters of mercury. Marston (38) used the sphygmomanometer for the distinguishing of emotional responses, particularly the differentiation of truth and falsehood. His work on the expression of emotion other than deceit seems for the most part inconclusive. Landis and Gullette (29) studied the changes in systolic blood-pressure which accompany emotional experiences. This investigation revealed no pattern of blood-pressure response to any situation or with any reported feeling or emotion. A single exception was noted in the case of the vascular response to the unexpected bang of a firecracker. In this situation a majority of the subjects gave a sharp rise in systolic blood-pressure followed by an equally rapid fall to the original level. Blatz (3) has reported a similar finding in choc situations. In a control experiment Landis and Gullette found that when blood-pressure measurements were taken at intervals of about 15 seconds, over a period of from one to four hours, during which the subject was passively reading material of a non-emotional nature, there were almost as great variations in blood-pressure as were found during the emotional situations. Landis, Gullette, and Jacobsen (30) made a statistical analysis of the correlation between the amount and variability of blood-pressure change with all other objective and subjective measures of emotional response which they had at hand for a group of 24 subjects. These correlations indicate that blood-pressure responses are negatively correlated with other emotional expressions.

Marston's work (38) has called a great deal of attention to the use of blood-pressure in the differential diagnosis of truth and falsehood. He claims that an increase of 8 mm. to 12 mm. of pressure in response to a question is sufficient to indicate that the subject is lying. The blood-pressure change during truth is that of a slowly decreasing general level. His method of determination was to take one or two readings of systolic pressure before a question was asked and again immediately after the question was answered, the difference in pressures giving the basis of diagnosis. Larson (35) has carried on this notion with a greatly refined technique. He used the Erlanger method for graphic recording of continuous blood-pressure, in prisons and in police courts, where the factor of truth or of falsehood was undoubtedly very real in the life of the subject. In general his results indicate that blood-pressure determinations can be used as a mild variety of the "third degree" for detecting falsehood. The technique is not as yet sufficiently understood so that one is safe in reporting an unqualified diagnosis in absence of other available supporting evidence.

---

[3] See Landis (24) and Landis and Gullette (29) for technical diagrams and descripitions of technique.

Brunswick (5) has made use of the Erlanger method of recording blood-pressure during a series of emotional situations in the laboratory. His work shows that blood-pressure does not furnish a differential pattern of emotional expression. Landis (25) has reported a study of blood-pressure changes by the Erlanger method in pronounced upset. No pattern of response was found other than the gradual decrease of pulse-pressure (systolic pressure minus diastolic pressure), which is comparable to the vascular change in "surgical shock."

In general, the findings from blood-pressure studies are as indeterminate as the findings of blood-volume changes when considered as expressions of emotion.

3) *Blood changes.* It has been suggested that the change in the pH of the blood might be considered as an expression of emotion. The normal pH of the blood is 7.2 as compared to 7.0 which represents a neutral solution; that is, there is a slight tendency for the blood to present an alkaline reaction. Certain investigators have stated that the blood changes from 7.2 to as much as 7.7 during an emotion. This change is usually accounted for as the result of the accumulation of carbon dioxide in the blood stream brought about either by the increased metabolism in emotion or as a factor in the maintenance of the level of emotionality in psychotic conditions. The studies of this acid-base ratio in the blood as expressions of emotion have not been overly successful. More significant results have been obtained by the use of the pH of saliva, which affords an indirect measure of changes in the blood. Starr (44) has reported that in emotionalized or psychopathic subjects there is an increase in the alkalinity of the saliva which is greater than the normal variations due to the physiological reactions of everyday life. Renshaw has been unable to confirm Starr's claims.

Another variable factor in the constitution of the blood is the presence of glycogen (blood sugar). Cannon has demonstrated that there is an increase of blood sugar during emotional stress. The kidneys constitute a mechanism which normally and automatically maintains a constant level of sugar concentration in the blood. When this level is exceeded, due to the liver excreting more sugar than the muscles use, the excess sugar appears in the urine. Studies made on both men and animals in periods of emotional stress demonstrate this glycosuria.

Mann (36) has shown that there is a change in the amount of sugar in the blood stream itself following the ingestion of a glucose meal. He presents curves showing the increase and decrease of this blood sugar in normal people and comparable curves in psychopathic patients who are suffering from definite emotional disturbances, e.g., acute depression, euphoria, mania, etc. These curves show a persistent glycemia in stuporous and melancholic patients whereas there is a hyperglycemia in the euphoric individual.

4) *Heart reactions.* There was a possibility that, even though the circulation fails to show a pattern of response, cardiac reactions, per se, might furnish some expressive pattern. There was the further possibility that the response of the heart might offer some clue to the balancing of activity between the divisions of the autonomic nervous system. In view of these possibilities Landis and Slight (31) made an electrocardiographic study of cardiac activity in response to choc stimuli. They found occasionally pronounced deviation from normal cardiac activity, but the appearance of this divergency was more closely related to the temporal correlation between the time of stimulation and the point in the cardiac cycle at which the stimulation took place than to any other factor. The results of this study failed either to confirm or deny assumptions concerning the activity of the autonomic nervous system. A study by Blatz (3) has given essentially the same findings. Blatz was particularly interested in the pulse-rate and found a "pattern" of pulse-rate in response to choc but no other change of importance.

Respiratory Reactions. While no theory basing emotion on respiration has gained any great amount of attention, the changes in the rate, depth, or pattern of respiration in response to emotional stimuli have long been noted and studied. Experimental work upon respiration in relation to affective experience is, on the whole rather unsatisfactory. Dumas (13), following Sikorsky, has tried to make a case for respiratory function, as diagnostic of mood in the psychotic individual. It seems possible that at certain times and in certain cases such diagnostic patterns might occur. However, prolonged studies of the same individual, who has been suffering from, say, acute depression, indicate that although there may be a slowing down of the respiratory rate there is no general pattern in his breathing. Miss Feleky (15) attempted to correlate various respiratory expressions, viz., respiratory work, respiratory volume, inspiration-expiration ratio, with facial expression and bodily activity during imaginary emotional experiences. Her findings are interesting but are based on one experiment on one subject for each "emotion." In view of these limitations it seems that little significance may be attached to the point of "pattern" upon which she insisted. Indeed most of her findings, when subjected to statistical analysis, show that the difference which she claimed is of no significance.

Landis (25) reports that breathing in pronounced upset shows marked respiratory work which slowly decreases; and, as the upset progresses, the respiration becomes more and more shallow. Blatz (3) found that, if choc stimulation occurs during expiration, the respiratory wave will at once be changed to inspiration; whereas if it occurs in inspiration, no change is noted. Skaggs confirms this while Landis and Slight (31) could not obtain such a change.

In the detection of falsehood, special use has been made of the inspiration-expiration ratio (I/E). Benussi (2) first investigated this function and claimed 100–per-cent certainty in diagnosis of the truth or falsity of the verbal response of a subject in highly artificial situations. Benussi claimed that the I/E ratio of the three or five breaths before telling the truth was greater than the I/E ratio of the three or five breaths after telling the truth. The ratio before lying was less than the ratio after lying. Benussi stated that it was possible to detect falsehood even when the subject was attempting to breathe regularly, keeping in time with the beating of a metronome. He found further that imagining a falsehood gave no significant breathing changes. Burtt (6) repeated, in part, Benussi's experiments, making use of an instrument which gave the I/E ratios directly. He found that he could diagnose falsehood from truth 20 per cent better by this method than would be possible by chance. (He further confirmed this by making use of a jury whose findings were no more accurate than one would expect to obtain by chance.) Landis and Wiley (32) attempted to determine the lower limit of actuality or reality in a deception situation which would give diagnostic respiratory or circulatory symptoms. Their situations were artificial, and no attempt was made to make the situation emotional or vivid. They found that respiration changes (I/E) were six per cent better than chance as a diagnostic symptom. No one has as yet been able to confirm Benussi's sweeping claims. It is evident that some factor is operating in the deception situation which does, to a certain extent, influence respiration. Landis and Wiley have advanced the theory that this change is a tendency to prolong expiration and to shorten inspiration, that is, a tendency to a gasping sort of breathing in deception but not in truth.

Considerable investigation has been directed toward the relationship between respiratory pattern or respiratory function and feeling (pleasantness or unpleasantness). Much of this has been done to check the tridimensional theory of Wundt. Drozynski (12) in an analysis of the I/E ratio in relation of pleasantness and unpleasantness found a tendency for the ratio to increase in pleasantness and excitement and to decrease in unpleasantness. Rehwoldt in a similar study confirms most of the findings of Drozynski. For the most part careful research has failed to bring evidence which does more than show that the respiratory functions are a variable by-product of affective experience and have nothing to do with the content of pleasantness or unpleasantness.

The findings concerning respiration as an expression of emotion may be summarized as follows: (1) Respiration does not vary directly with emotional excitement. (2) Certain respiratory patterns are diagnostic of certain situations. Choc and possibly deceit are situations for which certain experimental evidence is positive. (3) Simple pleasantness and unpleasantness probably are accompanied by respiratory changes which do not influence the content of the experience. (4) Respiration is undoubtedly a leading factor in emotion, but is not a differential expression of emotion.

Metabolic Rate. Metabolic rate refers to the sum of a series of energy changes within the organism. These changes may be either the accumulation of energy (anabolism) or the depletion of energy (catabolism) within the body. It is, of course, obvious that in emotional disturbance there is a marked change in the energy transformation and output. The earlier studies of the relationship between metabolism and affective experience were made upon the insane. The main notion of these experiments lay in the fact that an attempt was made to relate metabolic rate to the general mood or prevailing affective tone of the patient. The studies are conflicting so that the results do not indicate anything more startling than that the depressed or melancholic individual has a lowered metabolic rate, while the excited euphoric patient shows an increased rate. Graff (20) and his co-workers attempted to find the effect of emotion upon metabolism by use of hypnosis in which emotional reactions were suggested. They state that in one case they found as high as a 25–per-cent increase in oxygen consumption when the subject was told that he had an incurable cancer while the same subject showed only a 4–per-cent increase when he was told that he had been left a large sum of money. Their experiments lack control and verification.

Miss Totten (47) was the first to attempt to study metabolism, per se, as an expression of emotion. She recorded muscular movement, abdominal respiration, pulse-rate, and oxygen consumption on 14 subjects, subjected to a series of emotional situations. She found in six subjects an increase of from 5 to 20 per cent in oxygen consumption. In the remaining eight subjects there was no change worth noting. Her explanation for the increased metabolic ratings is that an increased volume and ventilation of the lungs due to the relaxation of bronchioles was caused by adrenalin.

The effect of pronounced upset upon the metabolism was the subject of an experiment by Landis. He made determinations of the basal metabolic rate daily for a period of three weeks, thus obtaining an average rate and the range of rates for each of three subjects. The subjects were required to go without food for 48 hours, without sleep for 36 hours, and were then subjected to intense electrical stimulation. At the end of this period it was found that during the time of anticipation before the stimulation there was a very marked increase in metabolism. This increase amounted to 16 per cent, 29 per cent, and 48 per cent, respectively, for the three subjects. When the actual stimulation was started, the metabolism fell off very rapidly and at the end of a six- or eight-minute period the metabolism had practically returned to normal. The metabolic rate during the hour after the stimulation was not markedly different from the normal rate, and the metabolism for the succeeding three or four days showed no deviation from normal. In an auxiliary experiment Landis found that the emotional experience aroused by listening to phonographic music did not cause any significant change in metabolic rate.

The metabolic rate is a very doubtful expression of any mood or temperament. Unpleasant or nocuous stimuli usually are accompanied by a transient increase in metabolism, while extreme upset is marked by a sharp increase followed by a decrease which is almost as rapid as the primary increase. The alterations in metabolic rate are not to be considered as expressions of emotion but as concomitants of emotion. There are several suggested causes for the increase in metabolism during emotion: (1) an increase in the secretion of adrenalin or of thyroxin with consequent release of glycogen and carbon-dioxide elimination; (2) an increased muscular tonicity or incomplete muscular relaxation of the voluntary muscles of the body; (3) relaxation of the bronchioles with consequent absorption of more oxygen; and (4) a greater metabolism of the central nervous system itself. The evidence for any one of these suggested causes is incomplete but suggestive. More experimental work is needed before one will be able to say with any degree of certainty which is the correct explanation or what is the proportion of the change which is due to any one of the factors.

Gastro-intestinal Activity. It is common knowledge that affective situations frequently lead to disturbances in gastro-intestinal activity. The nausea and disturbances of digestion accompanying upset have been known and remarked upon by psychologists for many years. James developed his famous theory of emotions on the hypothesis that the perception of the visceral changes constituted the differentia of emotion from other mental experience. His theory has led to numerous experiments.

The work of Pavlov has demonstrated that the mere presence of food in the mouth is sufficient to call out what he speaks of as "psychic secretion," even when no food has entered or can enter the stomach. It has also been shown that this secretion is altered by emotional stimulation. Cannon (8) found essentially the same facts for humans that Pavlov had demonstrated with animals. He notes that both the anger of an excited dog or the fear of the cat which is being barked at will inhibit salivary and gastric secretion. The activity of the smooth muscles composing the gastro-intestinal tract is markedly changed during affective experience. If an animal is given a bismuth meal and the progress of the food through the intestinal tract is followed by X-ray examinations, it may be demonstrated that unpleasant or nocuous stimuli lead to inhibition of the orderly process of digestion. It is also shown that intense excitement of a pleasant nature will inhibit digestion while mild pleasantness tends to facilitate activity.

The phenomenon of hunger, though not usually classed as emotion, fits excellently in practically any usual definition of emotion. A detailed consideration of these facts is given in another chapter of this book. The visceral reactions which are common to hunger are so similar to those which are produced in response to affective stimulation that the activity of the organism shows little that is differential in its expressions. It should also be pointed out that hunger has a specific organic concomitant which has not been found in any emotional expression.

Brunswick (5) has conducted a very interesting experiment upon the effect of emotional stimuli on the gastro-intestinal activity of the human being. He studied, by means of the gastric balloon, duodenal balloon, and rectal balloon, the gastro-intestinal activity during emotional situations. The situations were such as to bring about emotional disturbances in most of his ten subjects. He found that there was a loss or decrease of gastro-intestinal tone in fear, envy, disappointment, irritation, pain, and unpleasantness, while distress, surprise, and startledness usually increased tonus. Amusement, delight, admiration, appreciation, and disappointment never brought out gastro-intestinal change, while fear, relief, and envy gave contradictory reports. In all of his work it is obvious that specific changes in gastro-intestinal tonus are related to mild pleasantness or unpleasantness and to nothing more. He found no gastro-intestinal pattern reaction which would furnish a basis for the James theory of emotion. James' theory also depends upon the temporal relationship between the feeling and the visceral change. This change should take place, and subsequently the sensory qualities should be perceived and labeled as emotion or feeling. He found very contradictory evidence on this point. Some of his subjects showed the necessary temporal sequence, others failed to demonstrate it.

In an experiment carried on by Landis which has been mentioned above, the subjects went without food or sleep for two days, during which period records were taken of respiration, blood-pressure, gastric contractions, rectal contractions, and of metabolism. The stimulation at the end of the period produced emotional upset. The effect of this situation was to bring about a cessation of gastro-intestinal activity both of the stomach and the rectum. The effect of the stimulation upon blood-pressure, respiration, and metabolism has been described above. According to the presupposition of the James-Lange theory, there should be some pattern or order in the visceral and metabolic reactions in the upset of each subject since each of the three reported entirely different emotional titles for their experiences. The similarity of the reactions of the three and the haphazard physiological relationship may be best interpreted by assuming that there was a disturbance in the usual physiological integration; and that the reactions were merely those of random, chance occurrence, or of the selective balance of sympathetic nervous functions peculiar to each individual.

From all of the above studies of visceral reactions it will be noted that metabolism, gastro-intestinal activity, respiratory activity, and circulatory activity are all intimately interconnected. It seems that excitement, choc, or mild facilitation of activity creates a favorable balance so that visceral activity is facilitated. Nocuous, strong, or pronounced stimulation of any variety or "intense emotional experience" all tend to interfere with the complicated integration. The visceral responses are, to the best of our present knowledge, mediated over the autonomic nervous system or through the action of hormones acting directly on the organs. The disturbance in emotion lacks pattern, in the sense that pattern should be shown in the same way in all individuals. What is found is an individual pattern. A man experiences an emotion and every time he experiences it relatively the same reactions take place, but there can be no carrying over from one individual to another, or indeed from one period in the same man's life to some other period separated by months or years. The visceral changes give a basis for emotion but are not the emotion or constant expressions of it.

THE RÔLE OF THE ENDOCRINE GLANDS IN THE EXPRESSIONS OF EMOTION

Of recent years the relationship between the endocrine glands and emotion has been the subject of a great deal of experimental investigation. The interrelationship between the activity of the various glands of internal secretion and the autonomic nervous system is one of the most abstruse and complicated problems which faces the physiologist. It is a matter of biochemical and biophysical balance in which there are some eight or ten variables in relation to an autonomic nervous system of control, the fundamental action of which is only imperfectly understood. We may state in advance that certain internal secretions have a marked rôle in emotion. Others may be involved but we are unable to state specifically the effect of their activity. Still other internal secretions probably have no part in affective experience.

Suprarenal Glands. The work of Cannon (8) and his collaborators on the importance of the suprarenal glands in emotion constitutes a brilliant chapter in physiological psychology. A detailed discussion of Cannon's work is given in the chapter on "The Neuro-humoral Basis of Emotional Reactions." In order to complete the picture which we are endeavoring to give of the expressions of emotion, certain factors concerning suprarenal function should be pointed out.

Cannon has shown that the action of adrenalin (the internal secretion of the suprarenal glands) on voluntary muscle is on the motor end-plates of the nerves leading to the voluntary musculature. Adrenalin causes relaxation of certain of the smooth muscles and glands by direct chemical action upon the tissue itself. The effects of adrenalin may be summarized as follows: (1) increases the tremor in voluntary muscle; (2) causes relaxation of smooth muscle; (3) counteracts fatigue in voluntary muscle by affecting the myo-neural junction; (4) alters blood distribution in its inhibition of the sympathetic division of the autonomic nervous system and so changes the volume distribution of blood of the body; (5) alters blood-pressure by direct action on the cardiac muscle and in the contraction of the arteries; (6) hastens the rate of clotting of the blood; (7) relaxes the bronchioles; (8) causes the liver to release into the blood stream glycogen available for muscular work.

All of these physiological changes brought about by the action of adrenalin are, after a fashion, emergency reactions which enable an organism to meet a situation that will call for the quick and probably prolonged discharge of energy. No one of the reactions offers a pattern which could be considered as a differential expression of any of the "emotions," but the reactions are all of a sort which fit under our definition of emotion.

Marañon (37) has studied the effect of subcutaneous injection of adrenalin upon the human subject and reports that the subjects characterize the experience as a "cold emotion," that is, they are stirred up without the external reference or a situation usually associated with the experience. He also demonstrated that if one suggested an appropriate situation or experience during this physiological reaction, the course of the emotional reactions might be altered. For instance, if after administering adrenalin one suggested to the subject that his mother had suffered greatly on her deathbed, active crying and grief might result. Further research could be profitably carried out on this point.

Thyroid Gland. The thyroid gland is a ductless gland lying in front of the upper part of the trachea. Its internal secretion is known as thyroxin. Since we have no direct chemical or physiological test for the presence of thyroxin in the blood stream, most of our knowledge concerning its action comes from indirect or clinical evidence. There is no experimental work showing whether thyroxin is secreted in an emotional situation. It is probable that such secretion does take place since (1) emotion is expressed through the sympathetic nervous system, and the thyroid gland is innervated and stimulated through this same system; and (2) adrenalin stimulates both activity of the sympathetic nervous system and the thyroid gland. Although the experimental evidence is meager, the clinical evidence is rather clear. In hyperthyroidism or exophthalmic goiter we find a marked emotional instability in the patient. Such individuals over-react to all emergency situations or to nocuous stimuli and are in a state of heightened tension and irritability at all times. In hypothyroidism or myxedema we find a paucity of emotional expressions and responses. The patient is stolid, stuporous, and sluggish. Evidently the thyroid secretion furnishes a background for emotional response. That is to say, the level of the emotional reaction, the mood, the temperament, whether excited or depressed, is somehow the function of the presence of thyroxin in the blood stream.

The Pituitary Body. This is a small two-lobed body at the base of the brain. It is thought to give two internal secretions, tethelin and pituitrin. Uno (48) has shown that prolonged emotional stimulation in rats brings about changes in the size and chemical constitution of the pituitary gland. The exact nature of these changes is not clear. Our evidence from clinical medicine concerning the pituitary and emotion is meager. The action of pituitrin is somewhat similar to that of adrenalin and thyroxin while tethelin is probably concerned chiefly in bodily growth.

The Gonads. Gonads is a generic term applied to the sex glands. The internal secretions of these glands are called the interstitial secretions. The absence or deficiency of this internal secretion is quite marked and well understood, especially from studies which have been made upon eunuchs or eunuchoidal individuals. The hyperfunction is equally marked in the development and expression of sexual passion or lust, particularly in satyriasis or nymphomania.

Summary of Endocrine Action in Emotion. (1) The endocrine glands produce internal secretions which furnish the background or general tone of emotional experience. (2) In order that emotional experience may be considered normal, it is necessary that the interrelationship between all glands be well balanced. (3) There is no specific connection between any particular emotional experience or expression of emotion with any particular gland. (4) Adrenalin makes possible the excess output of energy in an emergency situation, while thyroxin furnishes the background persisting over a period of days or weeks. (5) Interstitial secretions are necessary for the adequate expression of sexual passion.

The Rôle of the Nervous System in the Expression of Emotion

The Peripheral Nervous System. The external expressions of emotion are under the direct control of the peripheral nervous system. In motor or sensory paralysis there is a marked interference with emotional expression. The lack of expressiveness in facial paralysis or in sensory anaesthesias of the face is well marked. There is, however, no evidence that the peripheral nervous system, as such, offers any other than the connecting linkage between central control and the expressions of the emotion.

The Spinal Cord. In decerebrate animal preparations, certain of the integrated reflex patterns are sometimes considered as expressions of emotion. The struggling or scratching which results from the painful pinching of the leg of a spinal dog is an example of this. In the human subject where the spinal cord has been severed we find exactly the same varieties of reflex coordination which might conceivably be considered as expressions of emotion, e.g., withdrawing from painful stimuli, automatic sex activity, etc.

The Mid-Brain and Thalamus. This region of the central nervous system is of particular and peculiar interest, since all evidence goes to show that it is here that the expressions of emotion are organized, integrated, and controlled. This area also acts as a relay station for all sensory impulses passing on to the cerebrum. It contains many of the nuclei which control vegetative activity of the body. The nucleus of the facial nerve is in this region. The neurology and physiology of this area has been discussed in the chapter on "The Neuro-humoral Basis of Emotional Reactions," *q.v.* For our purposes at present it is necessary only to show how this system acts in the organizing of emotional response. The integrity of the emotional life of the individual, the facility of expression, the range of emotional experience, are all somehow functions of the intimate connections of this old brain center. It is just possible that the predominance of facial expression in emotional experiences is due to the fact that the nucleus of the facial nerve lies in the thalamic region closely associated with other organization points for responses controlled over the autonomic nervous system. The presence of the control centers for respiration, circulation, digestion, electrical variations of the skin, sweating, secretion, and most of the other vegetative activities of the body which enter into any emotional experience are centered in this portion of the brain. The neural scheme of organization of emotional expression is somewhat as follows. Sensory nervous impulses passing upward on their course to the cerebrum relay in thalamus. The impulse may then pass in either of two major directions: (1) to the cerebrum where it enters into sensory, perceptual, and voluntary motor activity; or (2) part of the impulse may be diverted at the level of the thalamus and start neural activity from these centers. This activity constitutes what we have termed the emotional background of a reaction or reaction system.

The Cerebrum. For the purposes of the present discussion the cerebrum is an organ of voluntary and learned expression and of the inhibition or control of expression. All of the expressions of the face, the gestures, the voluntary bodily activity with which we convey knowingly our emotional reactions, are a function of the cerebrum. The question of the development and socialization of expression is one which falls more properly into the field of learning. We have rather arbitrarily considered that the true expression of emotion is an unlearned and more or less involuntary matter, since the treatment of expression as a form of learned response would lead us far afield from emotional expression.

The Autonomic Nervous System. The whole question of the function of the autonomic nervous system in the expression of emotion is one which has been treated in a previous chapter. Hence the treatment of the subject here will be very brief. All involuntary muscular reaction depends for its innervation upon the autonomic system. All organization, all pattern of response, all of the temporal relationships which may be considered as emotional expressions, are, therefore, somehow integrated by the autonomic system. There is, as has been shown by Cannon and his co-workers, a functional balance between the sympathetic division and the cranial and sacral divisions of this system. Such physiological pattern as is shown in the autonomic responses has been found in the balance between these divisions. Mild pleasure, or absence of pain, or strong stimulation is shown by a predominance of cranial and sacral function; while major unpleasantness, choc, upset, or violent emotion indicate activity of the sympathetic division. The entire mechanism is reflex in nature. It is so balanced that any portion of the visceral response which becomes predominant will quickly and effectively find compensation elsewhere in the system in all normal circumstances.

The Properties of Nervous Conduction. Lapicque (34) has developed a systematic explanation of neural activity based on the time-energy relations of the activity of functional tissue. He calls this peculiar property of living matter "chronaxie." Stimuli usually give rise to impulses which follow neural pathways in which the neurones are isochronous (have the same chronaxie). If the stimulation is strong or persists, the neural impulses which result will follow not only the isochronous pathways but will be forced into heterochronous paths. The taking over of nervous impulses from isochronous to heterochronous paths furnishes the neural basis for the experience of unpleasantness. The more intense the heterochronicity of the neural system, the more intense the unpleasantness of emotional experience. Pleasantness is to be regarded as the absence of unpleasantness.

Head has, on the basis of a very thorough and critical study of skin sensation, pointed out a distinction in cutaneous sensation quality which he terms protopathic and epicritic. Protopathic sensation refers to the undifferentiated and undifferentiating sensation quality from the skin which first appears on the regeneration of a divided sensory, peripheral nerve. This protopathic sensibility has the attribute of marked unpleasantness. Epicritic sensation is the acute, well-differentiated, and differentiating sensibility which marks normal cutaneous sensation. It has neither pleasantness nor unpleasantness save as they are elements of pain or lack of pain.

The problem of the nature of feeling (pleasantness or unpleasantness) is one which has long baffled physiologists and psychologists. The evidence which we have from Lapicque and Head represents only two of several theories which have been advanced as a physical basis for affective experience. A broader theory which is gaining some attention at present depends on the fact that the autonomic nervous system contains sensory as well as motor fibers. These sensory fibers carry impulses from smooth muscles and glands which are reported in mid-brain centers merely as coordinate or incoordinate activity. Coordination is then identified with perceptual pleasantness, and incoordination with unpleasantness.

The Electrical Expressions of Emotions[4]

The phenomenon usually termed "psychogalvanic reflex" was first called to the attention of the scientific world by Vigouroux, Féré, and Tarchanoff. Vigouroux pointed out in 1888 that the fact of variation in electrical properties of the human body had been studied and noted as early as 1786 by Bertholon. DuBois-Raymond's classical experiments on the electrical properties of living tissue and on certain of the electrical properties of the human body is basic to our knowledge of this subject. Vigouroux, in 1879, reported a study of variations in the resistance of the human body and mentions changes in resistance which are due to the *"sensibilité"* of the subject. Féré (1888) called specific attention to the fact that affective experience gave rise to variations in the resistance of the body. In 1890 Tarchanoff published a thoroughgoing study of the relationship between mental activity and the galvanic phenomenon of the skin. His work apparently grew out of duBois-Raymond's earlier investigation and was an attempt to establish the physiological basis for such electrical responses.

The notion advanced first by Féré that the electrical response was an expression of emotion has been of a great deal of interest to psychologists and has resulted in a long series of investigations. Veraguth, who named this phenomenon the psychogalvanic reflex, stated in 1907 that the electrical change was dependent on, or an indicator of *stress* in higher mental processes. Binswanger in 1907-8 seems to have been the first to hold that the reflex was specific to emotion. His work was part of the *Diagnostic Association Studies* of Jung and the Zurich Psychoanalytic school. He maintained that although the reflex occurs as a response to a variety of stimuli the mental process involved may always be grouped under the concept of emotion. Piéron in 1910 threw the weight of his influence to the support of the dogma that the reflex is specific to emotion. It is not clear from Piéron's earlier papers just how he derives this notion or what his evidence is. His later papers lack the unqualified assurance of his original statements. Peterson goes so far as to hold that the amount of galvanometric deflection is directly proportional to the strength of the emotion. Waller is responsible for persuading the majority of the scientific world interested in the phenomenon that it is specific to emotion. This he did largely through numerous demonstrations and papers given before scientific meetings in England, Belgium, and France. Wechsler is probably the only experimenter who has seriously investigated the problem of specificity of the reflex to emotion. Much of Wechsler's work does tend to confirm the hypothesis but he reports enough negative instances and findings which do not easily fit into the hypothesis to make one exceedingly skeptical of the fact that a thoroughgoing case can be made for the theory of emotional specificity.

---

[4] For a bibliography on this subject, see Landis and DeWick (28).

Radecki in 1911 and Abramowski in 1913 were the first to question sharply the assumption that the galvanic reflex is associated *only* with emotion. These investigators believed that the mental processes known as volition were more commonly associated with the reflex than were those of emotion. This notion has been followed up in the work of Aveling and his associates. They hold that the reflex is the correlate of conation and shows itself in affective experience only when the conative process is present. Other psychological investigators have held that the reflex follows every sensory stimulation, that it comes only in periods of stress, that it always accompanies the appearance of unconscious processes in behavior, that it is correlated with the complex oretic process, that it occurs in only purely ideational processes, and so on. Indeed a survey of the literature shows that psychologists have identified this electrical reaction with almost every conceivable mental process, and with only the most questionable of evidence to back their claims.

If we turn to the work of the anatomists, we find that Schilf and Schuberth demonstrated in 1922 that the electrical response occurred only when the autonomic nervous connections between effector and receptor were intact. They demonstrated that it was possible to sever all bodily connections between the two halves of the body of a frog, with the exception of autonomic fibers, and still to be able to elicit the reflex. They localized the control center for this response in the medulla. Foa and Perserico, using cats for experimental material, in 1923 followed the suggestion of Schilf and Schuberth. These experimenters found that the effector side of the reflex was localized in the autonomic fibers following peripheral motor nerves; that the brain control center of the reflex is localized in the corpora quadrigemina and in the apex of the calamus scriptorius; that peripheral stimulation starts impulses which traverse any sensory nerve, which in turn communicates with the bulbar region, giving rise to the response over the autonomic system. It seems quite clear from this anatomical research that the reflex is specifically controlled by the autonomic nervous system. This raises the question as to its physiological relationship to other autonomic responses.

It has been demonstrated by Gildemeister that the apparent change in resistance of the body is in reality a change in either the amount or the rate of polarization. He has also demonstrated that the apparent electrical output of the body is to be explained in terms of polarization. Since polarization is the important factor in the response, he has carried out very elaborate studies on the nature, rate, and factors affecting polarization in the skin and in living tissue. His results show that the alterations of polarization, usually called the psychogalvanic reflex, are dependent in the living organism upon the integrity of the autonomic nerve supply, and that the rate and duration of the reflex is a function of the sympathetic nervous control. Aveling and his co-workers have shown that there is a very close relationship between the appearance of vasoconstriction and the appearance of the electrical response. Darrow has shown that there is a one-to-one correlation between the appearance of the reflex and the secretion of sweat. An intimate correlation between the reflex and local skin temperature has been demonstrated. A moment's consideration will show that the vasoconstriction, secretion of sweat, and local skin temperature are all of them autonomic expressions.

To return now to the psychological claims concerning the response. The reflex is undoubtedly governed by the autonomic nervous system from centers in the mid-brain. In so far as emotion is organized and finds expression through mid-brain centers, the psychogalvanic reflex may be regarded as an expression of emotion. It is equally clear, however, that any other bodily or mental (?) activity may give rise to this same response. The use of the phenomenon in psychological work should be interpreted in the light of the functional activity of the autonomic nervous system. If autonomic responses are to be identified with emotion, and a reasonable case can be made for this assumption, then one would be correct in saying that the reflex is specific to emotion. The major objection to this claim is that it does violence to the usually accepted usage of the term emotion.

Emotion and Feelings

The term feeling has at least three accepted usages: (1) a term referring to the experience of pleasantness and unpleasantness, (2) a description of an experience which is less vivid than emotion, and (3) the report of vaguely perceived sensations which do not have clear perceptual quality. The differences in usage which are indicated by these classifications is a sufficient reason why the entire subject of feeling is so unsettled from the point of view of both the psychologist and the physiologist. One can never be quite sure whether the investigator is referring to the qualities of pleasantness and unpleasantness, to emotional experience, or to a sensation. We shall endeavor in this chapter to restrict our treatment of the subject to the facts of pleasantness, unpleasantness, and vaguely perceived emotion.

The older psychology, particularly that of Wundt and of Titchener, holds that feeling is one of the three basic elements of mental life. Feeling is marked by quality (pleasantness or unpleasantness), intensity, and duration, but lacks clearness. Wundt claimed that feeling was tridimensional in quality. These dimensions are tension-relaxation, excitement-depression, and pleasantness-unpleasantness. Most attempts to verify Wundt's hypothesis have been unsuccessful. It has been impossible to identify the tension, excitement, etc., with pulse, breathing, or any other physiological factor. Psychologically the Wundtian hypothesis has been tested by the method of paired comparison, i.e., the simultaneous presentation of two stimuli and the asking of the subject to report which is the most pleasant. This method has failed to give any marked support to Wundt's theory.

Physiological studies should, it seems, be able to furnish some clue to the nature of feeling or its relation to emotion. Following the lead of Lange, various experimenters have tried to associate pleasantness and unpleasantness with the volume distribution of blood within the body. The results of these experiments are very conflicting and have given us no real insight into the matter. Other theorists have held out the hypothesis that pleasantness and unpleasantness were to be identified with facilitation and inhibition in neural activity. Facilitation was supposed to carry pleasantness with it as an induced phenomenon; inhibition, to carry unpleasantness. Unfortunately there is no neurological substantiation of this theory, attractive as it is. Other psychologists have maintained that unpleasantness equals mild pain, while pleasantness equals the absence of pain. This is an attractive notion but one which has defied thus far either affirmation or negation. The work of Head (21) which has been cited above, on the protopathic and epicritic elements of sensation, is sometimes quoted as offering a basis for pleasantness and unpleasantness which might be drawn from the sensory qualities of experience. This hypothesis has certain elements which are extremely valuable in a physiological account of affective experience.

The work of Nafe (41) concerning the psychological or experiential basis of feeling has presented some very interesting points. His method is a variation from the usual introspective routine. In place of attempting to describe a feeling in dynamic terms, Nafe established the feeling process and would then cut it off short, after which he obtained as complete an introspective report as possible concerning the feeling. His subjects finally came to a description which identified pleasantness with "bright pressure." Phenomenologically this was described as kinaesthetic sensation from the upper chest and shoulder region. Unpleasantness was described as a "dull pressure" and associated with a pulling-down sensation in the region of the diaphragm. He concluded that sensory experience may be either affective or non-affective, but if pleasantness or unpleasantness were present there was an addition or complete change in the experience. Qualitatively, the affective increment was either a bright or a dull pressure. Pleasantness and unpleasantness were found inherently weak and mild though they have a certain pattern of density. Pleasantness came on gradually while unpleasantness appeared at full strength and at high intensity, passing over to emotion while pleasantness, as such, never did. Pleasantness and unpleasantness never appeared in isolation but always accompanied the sensory components although they were not organic sensations, nor the perception of organic sensations.

From the preceding samples of the literature certain conclusions concerning feeling as an expression of emotion may be drawn: (1) Feeling is certainly not a simple, easily identified, mental experience. (2) Feeling cannot be identified with any specific physiological occurrence or pattern of reactions. (3) Feeling may be either (*a*) the protopathic-epicritic element of sensation, (*b*) facilitation or inhibition in neural activity, (*c*) presence or absence of pain, (*d*) presence of reactions governed by the autonomic nervous system, or (*e*) the energy balance of the organism at any particular time. (4) Psychologically, feeling, in the sense of pleasantness or unpleasantness, appears as a unique experience but one which may be the product or by-product of many other experiential processes.

The expressions of feeling are more difficult to systematize than the expressions of emotion. This is, for the most part, due to the fact that feeling is less clear than emotion and that feeling is so much a matter of socialized response, whereas emotion is more nearly a native reaction. The expressions which are usually attributed to emotions should probably all of them be termed expressions of socialized feeling rather than of emotion. The entire subject of the development of feeling as a social reaction is one which lacks practically any experimental basis and one which is ripe and ready for attack by the experimental method. Developments in this field which are practically certain to occur within the next generation will profoundly modify all of our notions concerning psychology and the behavior of the individual in relation to the group.

Anhedonia

When James first announced his now famous theory of emotion, he stated that confirmatory evidence would probably be found in certain pathological conditions in which there was a decrease or lack of visceral sensation. This condition has been termed anhedonia. In 1905, D'Allonnes (10) described a case of anhedonia coupled with disorders of the time sense. This patient complained of an utter lack of the *feeling* of emotion while displaying all outward expressions of emotion. She suffered from complete pain, thermal, and visceral anaesthesia. Myerson (40) has reported a similar case of a young woman of cyclothymic temperament, who was suffering from "nervous prostration" as the result of an unrequited love affair. Following the affair she retained all the outward expressions of nervousness and emotion, e.g., crying, blushing, laughing, but protested that she had a complete lack of all "feeling." She reported that there was a total absence of hunger, thirst, fatigue, desire for sleep, joy, and fear. She would jump and scream at a sudden noise but denied all feeling of emotion. Each of these cases presents an interesting study of hysteria, and in the opinion of the writer should be considered as such. Other clinical case histories of total cutaneous loss of pain (analgesia) show no such lack of emotion or feeling. The explanation of the cases of anhedonia should be found in the dissociated functional habit systems, which are more easily discussed today in the terms of psychoanalysis than in the terms of physiological psychology.

Conclusions

The Relation between Emotion and Its Expressions. Many classifications and schematic diagrams of the interrelationship existing between the emotions or the expressions of emotion are to be found in the psychological and philosophical literature. In the present consideration of this subject the factor of classification of, or of the naming of, or an attempt to name the emotional expressions has been neglected. This has been done because it seemed that it was more worth while to adopt the experimental attitude, to study first the behavior, the physiology, the experimental literature on emotion rather than to begin by hanging tags on events not understood. It is doubtful that the appending of a classification would add particularly to the present treatment.

The descriptions which have been given of the expressions of emotion indicate that emotion, per se, is a unique physiological system of reactions for which the general boundary lines and landmarks have been determined, leaving the more intimate topography for future study. The expressions which we have studied, facial reactions, vascular reactions, respiration, gastro-intestinal reactions, etc., are all of them outstanding landmarks on our map. To carry the analogy farther, these reactions are the innate factors, the geographical or geological formations which are to be modified by the weathering action of environmental factors, thus more clearly defining those types of behavior which have been traditionally classified as emotions.

The question of nature versus nurture is as marked in emotion as in any other type of human behavior. It has been fashionable to speak of emotion as a natural reaction, one which is little varied by experience. Certainly this is a very inexact concept. What the natural emotional life of an indivdual might be like is an unknown territory. Emotional life is modified more rigorously in the growth and education of an individual than perhaps any other variety of human experience. The reason for the statement so frequently made that emotion is a natural reaction, unmodified by learning, is that emotional reactions occur in such large units of physiological disturbance as frequently to swamp the mental life of the individual.

There is no field in which there is a necessity of closer alliance between physiology and psychology than in that of emotion. The psychologist talks glibly about the effect of the endocrines or the autonomic nervous system on emotion while the physiologist is quite as careless in his remarks concerning the effect of the psyche, psychic life, or mind upon the factors which he calls emotional. Certainly a closer *liason* between investigators in this field would make the work of one more intelligible to the other and facilitate greatly our knowledge concerning these unique types of biological reactions which do play such an important rôle in human behavior.

The Expressions of Emotion. It is apparent that many of the expressions of emotion are innate. Some of the purposeful interpretations of expression are extremely interesting. The notion that many of the expressions of emotion are due to the survival of emergency reaction patterns is a most attractive one. The idea that many of the expressions are the survival of formerly useful habits is certainly as attractive. There are, however, many expressions which do not fall into any purposeful classification, as, for instance, laughter and crying. Another puzzling feature of this study is that of the way in which social factors modify or form expression. If, as many theorists have believed, certain expressions are mimetic of either past experiences of the individual or more particularly past experiences of the race, then we must have a type of neural functioning or an inheritance of modified nervous functioning which has not been hitherto described. It is safest to say that the major expressions of emotion are physiological survivals of earlier biological relationships within the species, and that for the most part the expressions of the adult are the result of learning or training.

The question of expression as a language mechanism is one which deserves some attention. It has been mentioned before that vocal reactions usually have emotional connotations which are not carried by the words spoken. The change in tones of the voice, the alteration in timbre, the occurrence of the vibrato, are all of them emotional expressions, some of which are learned, some of which appear to be native. The same is true for facial expression. The language of gesture which has been largely neglected in modern psychology certainly will add interesting sidelights to our knowledge of the subject. The entire question of "expression," in the sense in which it is used by the artist or the actor, is another which deserves the attention of the physiological psychologist. We know that certain gestures, certain bodily postures, certain varieties of reaction occurring in appropriate situations are, almost without exception, interpreted as emotional. Why? The answer is not easy. There is no experimental evidence. The answer, when given, will modify many of our present concepts concerning the nature of aesthetics and of the nature of socialized emotional expression.

Probable Lines of Future Development of Experimentation on Emotion. There have been pointed out throughout this chapter places in which further experimental evidence would be desirable. It remains only to add a brief list of those points which will probably be of importance in verifying or changing our entire viewpoint on this subject. The most important line of future research is that of the nature of the relation existing between emotion and learning, in the broadest sense of the terms. The acquisition of ability to interpret the expressions of emotion, the acquisition of facility of expression, the understanding of the probable underlying mechanisms of habit formation, will all of them be certain to modify our present notions. The second main line of future development will come most probably from the study of the physiology of the central nervous system and of the autonomic nervous system with relation to their action in the control, organization, integration, and modification of emotion. It has been stated in texts of psychology and of physiology for many years that the problems of emotion defied the experimental method. True, this statement has not always been made explicitly, but certainly this notion is implicit in practically all of our modern texts. It is to be hoped that the picture given here of the expressions of emotion or of the availability of emotion as an experimental field in both physiology and psychology may bring about a change in this attitude.

REFERENCES

1. Bard, P. A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *Amer. J. Physiol.,* 1928, **84,** 490-515.
2. Benussi, V. Die Atmungsymptome der Lüge. *Arch. f. d. ges. Psychol.,* 1914, **31,** 244-273.
3. Blatz, W. E. The cardiac, respiratory and electrical phenomena involved in the emotion of fear. *J. Exper. Psychol.,* 1925, **8,** 109-132.
4. Borgquist, A. Crying. *Amer. J. Psychol.,* 1906, **17,** 149-205.
5. Brunswick, D. The effects of emotional stimuli on the gastro-intestinal tone. *J. Comp. Psychol.,* 1924, **4,** 19-79, 225-287.
6. Burtt, H. E. The inspiration-expiration ratios during truth and falsehood. *J. Exper. Psychol.,* 1921, **4,** 1-23.
7. Buzby, D. E. The interpretation of facial expression. *Amer. J. Psychol.,* 1924, **35,** 602-604.
8. Cannon, W. B. Bodily changes in pain, hunger, fear and rage. (2nd. ed.) New York: Appleton, 1929. Pp. xvi+404.
9. Carr, H. Relation between emotion and its expression. *Psychol. Rev.,* 1917, **24,** 369-375.
10. D'Allonnes, G. R. La rôle des sensations dans les émotions et dans la perception de la durée. *Rev. phil.,* 1905, **60,** 592-623.
11. Darwin, C. Expressions of emotion in man and animals. New York: Appleton, 1873. Pp. vi+374.
12. Drozynski, L. Atmungs- und Pulssymptome rhythmischer Gefühle. *Psychol. Stud.,* 1911, **7,** 83-140.
13. Dumas, G. Traité de psychologie. Vol. I. Paris: Alcan, 1923. Pp. xiv +964.
14. Eng, H. The emotional life of the child. London: Oxford Univ. Press, 1925. Pp. vi+243.
15. Feleky, A. M. The expression of emotions. *Psychol. Rev.,* 1914, **21,** 33-41.
16. ————. The influence of emotions on respiration. *J. Exper. Psychol.,* 1916, **1,** 218-241.
17. Frappa, J. Les expressions de la physionomie humaine. Paris: Schmid, 1902. Pp. 140.

18. Gates, G. S. A test for ability to interpret facial expressions. *Psychol. Bull.*, 1925, **22**, 120.

19. ————. The rôle of the auditory element in the interpretation of emotion. *Psychol. Bull.*, 1927, **24**, 175.

20. Graff, E., & Mayer, L. Ueber den Einfluss der Affekte auf den Gesamtstoffwechsel. *Zsch. f. d. ges. Neur. u. Psychiat.*, 1923, **86**, 245-253.

21. Head, H. Studies in neurology. (2 vols.) London: Frowde & Hodder. 1920. Pp. 338, 537.

22. James, W. The principles of psychology. Vol. II. New York, 1890. Pp. vi+704.

23. Landis, C. Studies of emotional reactions. I. A preliminary study of facial expression. *J. Exper. Psychol.*, 1924, **7**, 325-341.

24. ————. Studies of emotional reactions. II. General behavior and facial expression. *J. Comp. Psychol.*, 1924, **4**, 447-509.

25. ————. Studies of emotional reactions. IV. Metabolic rate. *Amer. J. Physiol.*, 1925, **74**, 188-203.

26. ————. Studies of emotional reactions. V. Severe emotional upset. *J. Comp. Psychol.*, 1926, **6**, 221-242.

27. ————. The interpretation of facial expression in emotion. *J. Gen. Psychol.*, 1929, **2**, 59-72.

28. Landis, C., & DeWick, H. N. The electrical phenomena of the skin. *Psychol. Bull.*, 1929, **26**, 64-119.

29. Landis, C., & Gullette, R. Studies of emotional reactions. III. Systolic blood pressure and inspiration-expiration ratios. *J. Comp. Psychol.*, 1925, **5**, 221-253.

30. Landis, C., Gullette, R., & Jacobsen, C. Criteria of emotionality. *Ped. Sem.*, 1925, **32**, 209-234.

31. Landis, C., & Slight, D. Studies of emotional reactions. VI. Cardiac responses. *J. Gen. Psychol.*, 1929, **2**, 413-420.

32. Landis, C., & Wiley, L. E. Changes in blood pressure during deception. *J. Comp. Psychol.*, 1926, **6**, 1-19.

33. Langfeld, H. S. The judgment of emotion by the facial expression. *J. Abn. Psychol.*, 1918, **13**, 172-184.

34. Lapicque, L. Essai d'une nouvelle théorie physiologique de l'émotion. *J. de psychol.*, 1911, **8**, 1-8.

35. Larson, J. A. The cardio-pneumo-psychogram in deception. *J. Exper. Psychol.*, 1923, **6**, 420-454.

36. Mann, S. A. Blood sugar studies in mental disorders. *J. Ment. Sci.*, 1925, **71**, 284-305.

37. Marañon, G. Contribution à l'étude de l'action émotive de l'adrenaline. *Rev. fr. d'endocrinol.*, 1924, **2**, 301.

38. Marston, W. M. Systolic blood pressure symptoms of deception. *J. Exper. Psychol.*, 1917, **2**, 117-163.

39. ————. A theory of emotions and affection based upon systolic blood pressure studies. *Amer. J. Psychol.*, 1924, **35**, 469-506.

40. Myerson, A. A case of altered emotions bearing on the Lange-James theory. *J. Abn. Psychol.*, 1918, **13**, 239-249.

41. Nafe, J. P. An experimental study of the affective qualities. *Amer. J. Psychol.*, 1924, **35**, 507-544.

42. Seashore, C. E. Phonophotography as a new approach to the psychology of emotion. Feelings and emotions: the Wittenberg symposium. Worcester, Mass.: Clark Univ. Press, 1928. Pp. 206-211.

43. Sherman, M. The differentiation of emotional responses in infants. I & II. *J. Comp. Psychol.*, 1928, **7**, 265-284; 335-351.

44. Starr, H. E. The hydrogen ion concentration of mixed saliva considered as an index of fatigue and of emotional excitation, and applied to a study of the metabolic etiology of stammering. *Amer. J. Psychol.*, 1922, **33**, 394-418.

45. Titchener, E. B. Lectures on the elementary psychology of feeling and attention. New York: Macmillan, 1908. Pp. ix+404.

46. ————. Expression. The new international encyclopedia. Vol. VIII. (2nd ed.) New York: Dodd, Mead, 1920. Pp. 285-286.

47. Totten, E. Oxygen consumption during emotional stimulation. *Comp. Psychol. Monog.*, 1925, **3**, No. 13. Pp. 79.

48. Uno, T. Effect of general excitement and of fighting on some of the ductless glands of male albino rats. *Amer. J. Physiol.*, 1922, **61**, 203-214.

49. Watson, J. B. Behaviorism. New York: People's Instit. Publ. Co. (Norton), 1924-1925. Pp. 248.

CHAPTER 14

LEARNING: I. NERVOUS MECHANISMS IN LEARNING

K. S. Lashley
Behavior Research Fund, Chicago

The adult animal shows few reactions which have not been modified by experience. These modifications appear in a variety of forms whose relations to one another are not yet clear. Hysteresis, fatigue, adaptation or acclimatization, increase in excitability, progression of physiological states, associative memory or conditioned reflexes, after-imagery, immediate memory, and recognition have been variously distinguished as constituting different types of alteration. So little is known of the determining factors in any of these phenomena that they can be differentiated only in terms of the stimulus-response relationships involved. Hysteresis, fatigue, adaptation, and increase in excitability imply only a change in the intensity of stimulation necessary to elicit the response. In progression of physiological states (the successive reactions of *Stentor* reported by Jennings, the succession of movements in spinal irradiation described by Sherrington) the external stimulus is constant while the reaction varies. This type of alteration does not indicate any change in the innate pattern of internal connections between receptor and effector, but only a temporary change in the relative conductivity of different systems.

A different mechanism seems to be implied by the remaining types of modification. In associative memory a reaction is called out by a stimulus which, before training, was incapable of eliciting it. The reaction to one stimulus is transferred to another which differs from the first, not merely in intensity but in location on the receptor surfaces. This change in the relations of pattern between stimulus and response suggests a corresponding change in the pattern of the conducting system. It has led to the conception of the formation of new paths between receptors and effectors which is the basis of most theories of learning.

By many writers this is regarded as the type of all true learning and the effort is made to describe every sort of association in terms of combinations of simple associative reactions. Habits become chains of conditioned reflexes, images are motor reactions, logical memory the product of implicit trial and error. By others, reproductive memory is distinguished from habit or motor memory on the ground that the former does not involve activity of the effector systems, but is a purely intraneural process resulting in a reproduction of the sensory impressions orginally aroused by peripheral stimulation. Immediate memory differs from other forms in its temporary character and ease of inhibition. Logical memory and recognition are distinguished by Bergson and others, chiefly on the ground that they are not subject to the law of exercise, and made the basis for animistic speculation. Poppelreuter (33) considers all learning to be of the type of logical memory. Hering, Semon, Rignano, and others have held that memory is a basic property of protoplasm; neurologists have been inclined to regard it as the product of some particular organization of nerve-cells.

With this variety of description and interpretation we are justified in raising the question whether the concept of learning or of memory embraces a unitary process which can be studied as a single problem, or whether it may not instead cover a great variety of phenomena having no common organic basis. The question cannot be answered until the physiological mechanisms of a variety of habits and memories have been worked out.

Evidence as to the nature of learning and the organic changes involved is to be sought in a number of fields of investigation. The nervous system is of unquestioned importance for the adaptive responses of the adult organism, and a study of its development as well as of its mature structure is significant for an understanding of its functioning. The investigation of the factors which determine the differentiation and growth of the nervous system in the embryo bears directly on the problems of the origin of the functional connections for both instinctive and acquired behavior. Study of the limitations placed upon learning by injuries to the nervous systems of higher animals gives evidence on the same points.

Another means of attacking the problem of learning is by analysis of the sensory and motor components of habits and the general nature of the disturbances of memory resulting from destruction within the integrating system. This material sets certain problems of integration which any theory of the nature of the learning process must take into account. Finally, a study of the conditions under which learning occurs, of the agents which accelerate or retard the process and of the uniformity of their action provides the material upon which general theories of the nature of learning have been directly based. We shall take up these various approaches in the problem in order, then review the chief current theories of the mechanism of learning to determine in how far they are consistent with the facts revealed by experimental studies.

The Origin of Behavior Patterns during Growth

Most species of animals at birth show complex and well-integrated responses, for the learning of which there seems to have been little, if any opportunity. The neural patterns leading to these responses are a product of heredity and growth, in the same sense that other bodily structures are. The many recent attacks upon the conception of instinct have neither altered this fact nor advanced our understanding of it. The mechanisms contributing to the development of instinctive patterns are of interest, not only because a knowledge of them may contribute means for control of the fundamental patterns of behavior but also because the forces which contribute to development are probably not greatly different from those which enter into later functional activity and an analysis of them may throw some light upon the more difficult problems of integration.

Factors Influencing Growth. The fertilized ovum already shows certain structural differentiations: fertilization pore, centrosomes, nuclear structures, internal fibers, to which the succeeding stages of development may be referred. Before segmentation the axial direction is, in most cases, determined, and the first segmentation plane follows this polarization. The division into two cells at once introduces new complexities in the relation of parts. The contact of the two cells limits respiration through the contiguous membranes and introduces a secondary lateral polarity which is certainly influential in determining the position of the next segmentation plane. And so, throughout the history of segmentation, each new structure produced serves to complicate the physicochemical relations of the parts and lays the foundation for the next step in differentiation.

In part, the potential plasticity of the cells is retained so that when the influences of adjacent structures are altered they undergo a de-differentiation and develop into structures other than those which they would have formed had they retained their normal position within the organism. Thus in some frog larvae the ectoderm on any part of the body may be induced to differentiate into a lens under the influence of a developing eye-cup. But definite limits to the developmental potentialities of most tissues are set at an early stage of growth. Thus when differentiation of a limb bud is once begun, the growing tissue will form a limb, though it is transplanted to an entirely different part of the body. Growth thus presents a continuous compromise between this tendency to self-differentiation and a subjection of the tissues to influences and limitations of development arising from their relationships to other developing parts of the body. After the first few cell divisions, potentiality for development is probably never wholly unlimited except in the germinal tissue. On the other hand, the differentiated tissue, though limited to the formation of certain structures, is still subject to important influences from other parts. The developing limb bud from the right side will produce a right leg, even when transplanted to the left side, but if its growing point is injured it may produce two legs, a right and a left, the polarity of each determined by the near presence of the other.

The forces contributing to the development of the nervous system are essentially the same as those entering into the development of other systems. Neural tissue, once differentiated in the neural tube, grows into nerve-cells, in large measure independent of its environment, and normal production of gross structures, such as optic vesicles, cerebellum, or mid-brain, continues even when the rudiments of these structures have early been transplanted to unusual positions. But although self-differentiation of such gross structures appears and the general characters of nervous tissue are irreversibly determined early in development, at the time of constriction of the neural tube, the fate of the individual neuroblasts (the undeveloped nerve-cells) is determined much later and primarily by their relations to other structures. This is illustrated by the experiments of Hooker and of Detwiler (8) upon the effects of reversing segments of the spinal cord of salamanders. In the normal cord the first coordinating axons developed from the neuroblasts grow toward the head. If a portion of the neural tube before the development of axons is cut out and replaced so that its normal anterior end is now posterior, the first axons formed are still directed toward the head. They thus grow in the opposite direction, with respect to the position of the neuroblasts from which they are derived, to that which they would normally have taken. The course of the fibers is determined, not by the character of the cells from which they are derived, nor by the immediate surroundings of those cells, but by their relation to the polarization of the whole body.

Child (7) points out the importance of the rapid growth of the posterior segments of the embryo at this time in establishing a gradient and determining the anterior direction of growth of the neurons, and the later suppression of these primitive nervous connections and the establishment of others when the activity of growth in the posterior region declines.

In determing the direction of development of the nerve-fibers, mechanical, chemical, and probably electrical stimuli play a part. Harrison has shown that neuroblasts growing in tissue cultures are responsive to mechanical and chemical stimuli. Recent theories of the factors in growth have inclined more to stress the electrical or electrochemical. Kappers (20) in the doctrine of neurobiotaxis has developed a general theory to account for the gross structural relationships and the finer connections of nerve-fibers. He assumes that excitatory processes give rise to differences in electrical potential, resulting in currents to which the neuroblasts are susceptible. Under the influence of these electric currents the axons are directed and grow toward the cathode, the dendrites toward the anode. When the neurons begin to conduct, the passage of the nerve-impulse gives rise to differences in potential which serve as further stimuli to cell growth and determine the intercellular connections which are established. Kappers' theory of selective interconnection of cells assumes that when a cell is stimulated its conductivity is increased. If, then, another cell in its vicinity is excited, the resultant potential will discharge along the previously excited neuron, the dendrites of which will be attracted toward the new source of excitation. The axon will at the same time grow out along the path of electric discharge, until contact is made between it and the attracted dendrites. In this formulation Kappers has probably over-emphasized the rôle of intraneural conduction at the expense of more general organic forces. Carmichael (4) has shown that the mechanisms of the first reflex responses of salamanders develop normally although the embryos are continually immobilized by a general anaesthetic. The doctrine that the reflex paths are established as a result of sensory stimulation begs the question in any case, for, in assuming that electrical currents from the point of stimulation determine the direction of growth, it presupposes the existence of those paths of high conductivity between sensory and motor zones whose origin it seeks to explain.

The broader principles of polarity, of physical, chemical, and electrical interaction among the parts of the developing embryo sketched by Child seem a more fundamental approach to the problem of morphogenesis, although they do not supply an explicit theory of the selective connection of neurons which is supposed to underly the reflex response.

This theory of specific intercellular connections as a basis of response, so widely accepted in psychological and physiological writings, calls for closer scrutiny. Must we assume that continuous and relatively isolated paths from specific sense-organs to effectors are laid down in growth, or are the anatomical details of growth of relatively little importance for the later functional relationships? The enormous variability in details of cellular structure gives little support to the strict anatomical interpretation of the problem. The evidence for functional plasticity accords with it even less. Detwiler (9) has shown that when the limb buds of salamanders are transplanted to positions behind their normal ones they are, at least in part, innervated by the outgrowth of fibers from neuroblasts which would normally supply the intercostal muscles. Nevertheless, when the connections are established, the movements of the limb exhibit normal coordination. Still more striking is the case in which injury to the growing tip results in the production of two oppositely symmetrical limbs from the same bud. In this case the coordination in movements of the limbs is determined by their symmetry and not by their anatomical position. Of course it may be assumed for such cases that each developing axon seeks out the muscle which its central connections fit it to innervate, but there is evidence from experiments on nerve-suturing that axons which originally supply one set of muscles will readily grow to and innervate another, so this assumption of selectivity in growth has little support.

In a later section we shall present other evidence against the importance of the individual neuron for behavior. The evidence from growth is not conclusive on this point, yet it requires rather far-fetched assumptions concerning the selective attraction of neurons for developing axons to save the theory of specific cell-to-cell connection in the face of such experiments as those of Detwiler (9) and Weiss (41). The latter has recognized the difficulty and abandoned the theory of specific paths in favor of one based upon the differentiation of muscle-fibers to respond to critical frequencies of excitation derived from any source.

After birth, in mammals there is a protracted development of new patterns and potentialities for behavior. These have been roughly classed as maturation of instincts, the appearance of delayed instincts, learning, and the growth of intelligence. There is little cell division in the nervous tissue after birth, but transformation of neuroblasts into neurons continues for a long time. The basis for further development of behavior is to be sought in this and in further growth and proliferation of nerve-fibers, myelinization of fibers, increase in cell size, or chemical changes which increase functional efficiency. We have little basis for judging which of these is the most important in determining patterns of behavior.

Maturation. Breed, Bird (2), and others have given evidence that without opportunity for practice there is improvement in the functions appearing at birth. Whether this is due to the development of further specific integrations or to a more general improvement in coordination is unsettled. Bird's observations suggest that the latter is probably the important factor. Tilney and Castlemajor report a close correspondence between the development of myelin sheaths in the tracts of the medulla and the appearance of new patterns of reaction in kittens.

Delayed Instincts. The most conspicuous group of delayed instincts are the activities which appear at the time of sexual maturity. There can be no doubt that many of these are complex patterns of behavior which appear fully integrated during the sexual cycle, without previous practice. Mating, nest-building, recognition and care of the young, are all activities of this sort. The sex behavior of the male rat has been most thoroughly studied. Steinach and, more recently, Stone (35) have shown that the internal secretion of the testis is usually essential for the first appearance of the behavior. In some way the particular neural pattern is activated by the hormone so that the mating response is given to stimuli which are otherwise inadequate. We can imagine several possible ways in which this might be brought about. The effect of the hormone might be a mere increase in vigor resulting in a general increase in nervous excitability. There might be an action upon the vegetative nervous system to increase the tone of the pelvic organs (for which there is some experimental evidence), and this, in accord with Kempf's theory of segmental strivings, might reflexly facilitate the spinal integrations; these possibilities are opposed by evidence of the specific effects of the male and female hormones. Finally the hormone may act selectively upon certain organized systems within the brain or cord, either to stimulate growth or to increase their excitability. The possibility of stimulation of growth seems ruled out by the rapidity with which the reactions are restored in castrates by implantation of testicular tissue and also by the fact that, in other animals at least, the reactions may sometimes appear before sexual maturity.

The most probable explanation of the delayed sexual reactions thus seems to lie in the special sensitivity of a definite neural system, previously organized by growth, to a specific chemical circulating in the blood. Of course there are many complicating factors. The male tern reacts at once to an egg in his nest, the female only after she herself has laid an egg; the feminized male rat will respond to young, the female only after she herself has borne a litter. The effectiveness of the hormones is modified by many factors. Habits soon come to play a rôle and obscure the congenital integrations.

The experimental studies indicate the importance of three factors for the initiation of delayed instincts: the growth of fiber connections, the maturation of these fibers as indicated by the formation of the myelin sheaths, and the chemical activation of neural systems. There are probably still other factors involved, as yet unrevealed by actual experiments.

The Growth of Intelligence. The increase in potentialities for complex behavior which parallels physiological development may, like the maturation of instinct, be ascribed to any of a number of changes in the central nervous system. The studies of Hammarberg and of Bolton show that there is a measurably smaller number of mature ganglion cells in the cerebral cortex of the imbecile than in that of the normal individual of the same age. Data which we shall present in a later section indicate that the complexity of the habits which can be formed and functional efficiency in performance is conditioned by the mere mass of cerebral tissue. These facts suggest that the development of intelligence with age may be a function of the increase in number of mature ganglion cells resulting from the differentiation of neuroblasts. There is a little evidence that the development of nerve-cells is stimulated by functional activity so that it is possible that education as well as innately determined growth processes contributes to the growth of intelligence. The independent development of neuroblasts, on the other hand, is difficult to harmonize with the constancy of the intelligence quotient, since the latter would require a great regularity in the number of neuroblasts reaching maturity each year. The alternative, that the grade of intelligence is determined by the number of neurons formed at birth or shortly after, and that growth of intelligence is the result of a gradual maturing of these cells, is not ruled out by any available evidence. The work of Travis and Hunter (38), showing a relation between the rate of reflex conduction and intelligence, may be interpreted as evidence that physiological maturation rather than increase in number of functional neurons underlies the growth of intelligence, yet the possibility that reflex conduction time may be influenced by the mass of facilitating impulses from higher centers and thus by the number of mature cells is not excluded.

Even before the development of the neuroblasts the organism presents a complicated pattern in which, although there are no known specific coordinating structures, a great deal of integrated activity of the parts is carried out, slowly to be sure, but with great precision. The coordinating processes are chemical diffusion, the creation and conduction of electric currents, and a high degree of selective sensitivity to these factors on the part of the developing cells. With maturity these forces do not cease to exist and it would be strange if the susceptibility of the nerve-cells to their action were to terminate with the completion of their growth. The myelin sheaths may serve to remove the axons from a large part of such influences, but the unmedullated portions of the cells within the gray matter still remain exposed and subject to the general chemical and electrical influences as well as to specific excitations from other neurons. We have as yet no means of estimating the part played by such extraneuronic influences in determining coordination. There is little direct evidence of their action, yet the difficulties which the reflex theories encounter in application to the more complex forms of behavior may force us to seek the sources of integration in such factors rather than in specific connections of neurons.

Development of Learning Capacity in the Evolutionary Series

Few animals lack tissue specifically differentiated for the conduction of impulses to movement. If we are to study the capacities for modification without such structures, we must turn to plants or to the rhizopods among animals. The various modifications of behavior reported for plants may be grouped roughly in three classes: acclimatization, adaptation to stimuli, and the acquisition and maintenance of various rhythms of activity. The first of these shows few analogies with learning in animals, since it seems to involve an alteration of the tissues through the local action of environmental agents upon them. Adaptation to stimulation has been described most adequately for *Mimosa*. When these plants are stimulated by mechanical or other nocuous stimuli, they respond by folding their leaves. With repeated mechanical stimulation it becomes more difficult to elicit this response, and plants may finally be brought to a state in which they will fail to react to continuous strong stimulation. Left unstimulated for a time, they recover the capacity to respond.

Alterations in rhythmic activities have been reported for the time of discharge of spores in several marine algae and for the growth processes in relation to illumination in higher plants. Their relation to other phenomena of adaptation is not clear.

Acclimatization in *Amoeba* has been described by several writers, but it is not clear that anything beyond a general change in excitability is involved. Mast and Pusch have recorded changes in behavior of *Amoeba proteus* which they interpret as analogous to learning in higher forms. The *Amoebae* were repeatedly stimulated by a beam of light directed across their paths, and the number of contacts with the beam before a complete reversal of direction occurred was recorded. This was repeated 27 times with each of five animals. On the average there was a gradual reduction in the number of false starts before the reversal of direction, which suggests a learning curve. The animals were kept on vaseline-sealed slides in darkness except when tested under the microscope. No controls were reported to rule out the influence of the continued adaptation to darkness or the effects of possible progressive chemical change in the medium. Only two of the five individual records show progressive improvement and there is no indication of loss of the effect after interruptions of training as long as 50 hours. The data are thus inadequate to establish an identity of the results with associative memory.

The problem of modifiability of behavior in other protozoa is complicated by the presence of a primitive nervous system in the form of a neuromotor apparatus, which is present in the flagellates, infusoria, and probably in other higher forms. The function of this apparatus for conduction and coordination of movements has been demonstrated, but little is known as to the nature of its activities beyond what may be inferred from the observed coordinations of ciliary beat such as the progression of waves of movement, reversal of direction of such waves in restricted areas, and the like.

Adaptation to mechanical stimulation may readily be observed in sessile forms such as *Vorticella* and *Stentor*. Summation of the effects of continued stimulation leading through a series of avoiding reactions to a final release from the substratum was described by Jennings for *Stentor*. Attempts to demonstrate effects of practice by Day and Bentley, Metalnikow, Buytendijk, and others have given questionable or negative results.

The available evidence shows for organisms lacking a true nervous system the possibility of decreased excitability following continued stimulation, the summation of stimulation resulting in increased excitability under certain conditions, and something like an irradiation of the effects of stimulation, resulting in a succession of responses, but there is no demonstration of an acquired change in the pattern of response or of any phenomena analogous to the associative learning of higher forms.

Among metazoa the evolution of the nervous system has involved three radical steps: the transition from syncytial to synaptic organization, the transfer from the ventral position of the nervous system in invertebrates to the dorsal position in vertebrates, and cephalization or the concentration and development of nervous tissue in the head region. We shall inquire what influence each of these steps has had upon the capacity for modification of behavior.

In the coelenterates and echinoderms, coordination of movement is probably brought about by a diffuse irradiation of impulses from the areas of stimulation through a continuous network of nerve-fibers. Even in the simplest of these forms there is some concentration of nervous tissue in the oral region and at the bases of the tentacles, and the spread of excitation seems most rapid through these regions of denser tissue, so that very specific reactions may be given to localized stimulation. In the more highly differentiated forms, anemones and many of the echinoderms, there are well-marked conduction paths forming a nerve ring about the mouth region, with radiating nerve-trunks to the organs of the radial segments. In the sea-cucumbers (*Holothuria*) a synaptic type of organization probably makes its first appearance. Behavior of forms with a simple nerve net has not been correlated in detail with nervous function. The behavior is quite complex and implies an elaborate interplay of diverse impulses from different excited regions, modified in various ways by the general level of physiological activity.

In the simpler coelenterates (*Hydra*) adaptation to stimulation and summation of the effects of stimulation have been demonstrated. Successive patterns of behavior may be called out by continued stimulation, as in *Stentor*, but there is no evidence that such progressions can be modified by practice. In the anemones Jennings found changes in behavior depending upon the state of hunger, upon acclimatization to stimuli, and upon the previous reactions of the animals. Of the latter the most striking is the maintenance of postures imposed by the irregularities of the surfaces upon which the animals live. The posture is maintained by a local contraction of one side of the polyp which is not relaxed during the usual minor changes in form. If the polyp is forced to a maximal contraction by strong stimulation, the bend does not reappear when the polyp again extends. Apparently the enforced position results in an area of maximal tonus within the nerve net, which persists until it is suppressed by a general maximal excitation. Jennings suggests that this may represent in rudimentary form the mechanism by which habits are retained in higher animals.

It is not certain where the synaptic type of nervous system first appears in the evolutionary series. Crozier, on grounds of behavior, ascribes it to echinoderms, and it is in these forms that the first evidence for reorganization of reflex connections as a result of practice appears. Experiments with starfish have been reported by Jennings (18) and by Ven. Usually two of the rays of the starfish, those adjacent to the madreporic plate, are more active than the others, take the lead when the starfish crawls, and play the chief part in the righting reaction. Jennings restrained the movements of these rays, forcing the use of others for righting the animals when placed on the dorsal surface. Training in the use of rays previously not used in the righting reaction for from 70 to 180 trials was followed in some cases by a more frequent use of those rays than of any other, and effects of the training persisted for at least five days after the termination of the training period. Ven restrained the two more active arms of the starfish by wire wickets, forcing them to escape by crawling with the less active arms in advance. He found a gradual reduction in the time required for escape.

The results of these experiments may be interpreted as indicating a shift in the lead or dominance from one region of the nerve net to another. There is little evidence that more precise changes in the pattern of coordinations can be established. When two rays were forced by training to take the lead in activity in Jennings' experiments, they might be so twisted as to oppose each other and make the righting reaction very difficult. Training failed to alter the tendencies to such opposed movements.

The value of the anatomical correlations here is somewhat lessened by the fact that the actual structures responsible for the synaptic type of conduction (transmission of the nerve-impulse in one direction only) is not known. The coelenterates and echinoderms give evidence in their behavior of a predominance of diffuse conduction in either direction, which does not permit of precise coordination of limited structures such as is involved in associative reactions. If they showed precise conditioned reflexes we should probably conclude, not that this type of learning is possible in a nerve net, but that the structures characteristic of the intercellular connections of higher forms are not the essential ones for synaptic function.

In animals in which a definite synaptic nervous structure is easier to demonstrate, the association of reactions with stimuli which were not sufficient to elicit them before training is readily obtained. Yerkes (42) established in a manure worm a habit of turning to the right in a simple maze. Indications of learning appeared after 20 trials, although absolute consistency of performance was not established within 1000 trials. Removal of the first five segments of the body, including the cerebral ganglion resulted in only temporary and slight disturbance of the habit. Regeneration of the cerebral ganglion, on the contrary, was followed by a loss of the habit. In addition to the usual conditions of maze learning, the experiment offered an opportunity for the formation of a simple sensory association. The entrance of the cul-de-sac was paved with sandpaper, beyond which the punishment grill was located. The animal might thus learn to associate the mechanical stimulus of the sandpaper with the avoiding reaction to the electric current. Inspection of the tables shows, however, that the ratio of contacts with the sandpaper to contacts with the grill remained practically constant, within its probable error, throughout the thousand trials of training so that there is no indication that this association was formed.

Broad generalizations are not justified by this one experiment but it suggests that the learning is a function of the entire ganglionic chain rather than of a limited part, and that the activities of the regenerated anterior ganglia were sufficient to disrupt completely the acquired activity of the posterior segments.

The arthropods and the higher mollusks approach the lower vertebrates in the complexity of their reactions and the readiness with which habits may be acquired. A wealth of experiments shows this and we need cite only a few. Turner's studies (39) show conditioned reflexes to sound and light in moths and roaches. The rate of learning does not seem significantly slower than that of lower vertebrates. The extensive experiments on form and color discrimination of bees by von Frisch show the ready formation of habits involving rather fine discriminations. Goldsmith reports the establishment of a discrimination habit in the octopus. The shelled mollusks seem to form associations less readily (Thompson).

Evolution of associative memory may be assumed to involve an increase in the rate at which association may be formed; increase in the complexity of associations, that is, in the number of elements or related processes which may be subsumed under a single habit, and perhaps qualitative changes such as are implied in the distinction between habit and reproductive memory. So far as the available studies may be taken as final, there is little difference between the higher invertebrates and the lower vertebrates in any of these respects.

The Rate of Learning. Goldfish require a greater amount of practice to learn a simple maze than do roaches to learn a more complicated one. The discrimination habits described for insects equal in complexity any reactions which have been reported for the cold-blooded vertebrates. It is thus clear that the shift from the ventral to the dorsal position of the main nerve-trunks has not, in itself, entailed any fundamental change in the capacity for learning.

Nervous evolution in the vertebrates has involved increase in the relative size of the brain and body, the development of the cerebral cortex, and the differentiation of the so-called association areas as the chief alterations which might be expected to change the capacity for learning. Speed of conduction and probably of all other physiological processes is increased, but the importance of these factors alone for learning may be largely discounted by comparison between birds and mammals and between various orders of mammals.

Comparable quantitative studies of the rate of formation of simple associations or the limits of training in the different classes of the vertebrates are rare so that there are few data upon which to base comparisons of function with structure.

There is little evidence that the rate of formation of simple mechanical habits changes in the mammalian series. Pavlov (29) reports that the conditioned salivary reflex may be formed in about 60 combined stimulations in the dog. Gley and Mendelsohn did not establish it in man in 40 trials, and Lashley failed after a much larger number of combined stimulations. Conditioned motor reactions seem to be formed about as readily in the dog as in man when the discriminations called for are relatively simple.

Pechstein (30) trained rats and human subjects in mazes having identical patterns. In one set of experiments the subjects were required to learn the maze by sections ("part method"), in another to learn it as a whole. The results were the following.

	TRIALS		ERRORS	
	Whole	Part	Whole	Part
Rats	27	30	217	199
Humans	12	23	126	237

Considering the far greater novelty of the entire training situation for the rats than for the human subjects, the results do not indicate any significantly greater learning ability in the latter.

Such data are not conclusive, but do suggest that the rate of formation of simple habits has increased little, if at all, through the evolution of the cerebral cortex. Other facts bear out this opinion. Simple habits are acquired by the feeble-minded about as readily as by normals. Extensive cerebral lesions markedly retard the formation of complex habits, yet produce little effect upon the rate at which simpler ones are formed. In opposition to these facts it may be urged that they concern only acquisition through trial and error in which chance plays a large part, and that direct association, as in immediate and incidental memory, is much more rapid in man than in lower forms. It is difficult to get objective evidence on this point, but there are many indications that even the rat, in the learning of mazes and problem boxes, forms a large number of immediate associations which are comparable only with incidental memories in man. Thus, a nearly decerebrate rat will often re-enter a cul-de-sac more than 100 times before passing on to the next. The normal animal rarely re-enters a cul-de-sac, even on the first trial in the maze, without intervening exploration of other parts. With no differences in bodily orientation to account for the diverse behavior, we seem forced to conclude that the entry into a cul-de-sac leaves traces in the normal rat which for a time inhibit re-entry.

Limits of Training. So far as one may judge from existing evidence, the higher invertebrates compare very favorably with the lower vertebrates, even with the lower mammals, not only in the rate of formation of simple habits but in the complexity of the habits which may be acquired. There are no comparable studies including widely different types of behavior, but certainly the recognition of the nest locality from visual cues reported for honeybees, exceeds in complexity any habit which has been observed in fishes or in the smaller rodents.

Within the vertebrate series the limits of training seem to parallel very closely the degree of development of the brain and particularly of the cerebral cortex. The habits reported for rodents, carnivora, lower and higher apes, both under experimental conditions and in the field, form a definitely graded series with respect to the number of elements and the complexity of the relationships involved. It seems also clear that the limits are set by the character of the central organization rather than by the sensory or motor capacities of the animals. Thus it is practically impossible to establish a habit of visual pattern discrimination in the rat with the conventional Yerkes discrimination box. We have many times trained animals with a vertical and a horizontal line as stimuli for 500 to 800 trials without clear evidence of discriminations. The same animals will distinguish lines immediately, if the lines are objects to which they must jump. The rhesus monkeys have a motor and sensory equipment not sensibly inferior to that of the chimpanzee. The habits in which they have been trained are, however, far simpler.

The comparative study of learning in different animals gives little evidence that evolution has brought any change in the rate of formation of the simpler habits. On the other hand, there is a fairly consistent rise in the limits of training and in the formation of complex habits with ascent in the phylogenetic scale. This raises the questions of the meaning of complexity in learning and its relation to the problem of insight or intelligence. Does increase in the number of similar elements set a limit to learning; that is, may complexity of habit be stated as a function of the number of single associations required, or is a qualitative change in the nature of the relationships among the elements the important factor in determining the limits of training? The data at hand are not easy to interpret. In studies of human learning, results have been obtained ranging from those of Henmon, showing practice proportional to the length of the series, to those of Foucault, showing practice proportional to the square of the length. The majority of experimenters have found that the difficulty is disproportionately increased with length even for meaningful material. This would indicate that mere multiplication of elements might set a limit to learning. In tests with rats we have found that increasing the number of culs-de-sac in the maze does not proportionately increase the practice required for learning, if the animals are normal, but does in animals with extensive cerebral lesions.

Even in the case of maze learning by defective animals or the memorizing of digits by normal men it is not clear that the difficulty of learning large complexes is due to the number of associations to be formed rather than to the inability to see significant relations within the complex. This problem is still unsolved, and our ignorance of it permits the claim, on the one hand, that all habit or memory is merely the linking together of conditioned reflexes and, on the other, the assertion that learning is always a product of insight.

Learning in Relation to Neural Structure

The Quantity of Tissue. Without much quantitative evidence, one is led to the view by a general survey of learning in the mammalian series, that the size of the brain, both absolute and relative to body weight, plays an important part in determining the limits of training. Lapicque (23) and others have computed the ratio of brain and body weights in a number of animals and sought a general formula to express the amount of increase in brain weight for unit increase in body weight. Lapicque concludes that between different species of animals and between the sexes in man this ratio is expressed by the formula $k = \frac{E}{P^{0.56}}$, where k is a constant, E is brain weight, and P is body weight. That is, the weight of the brain increases as the square root of the body weight. Within a given species the exponential relation is reduced from 0.56 to 0.20. Other factors than body weight contribute to the relationship. The most important is the surface area of the sense-organs, which determines the number of cells in the corresponding sensory regions of the nervous system. With due allowance for this, Lapicque is inclined to think that the above equations express the requirements of nervous tissue for the average intellectual level of the groups in question and that any marked departure from the ratio determined for a given group indicates an intelligence above or below the average for the group. There are no valid data on behavior to serve as a check on this conclusion.

Studies of the effects of cerebral destruction in animals show a definite relationship between the quantity of functional tissue and the capacity to learn. Lashley (24) found a correlation of 0.86 ± 0.03 between the extent of brain injury and the number of errors made by rats in learning a maze. Other experiments indicated that retention is also a function of the amount of intact tissue. The observations of Hammarberg and of Bolton indicate the importance of the number of ganglion cells for learning in man, and, although there are no adequate quantitative data, the opinion is frequently expressed that the degree of deterioration in man after lesions to the cerebral hemispheres is related to the extent of destruction.

Relation of Different Structures to Learning. Many writers of the last century ascribed learning wholly to the cerebral hemispheres. The demonstration of habits in invertebrates and fishes disproves this opinion, but still leaves the question whether the developed cerebrum has taken over the learning function from lower centers. Trevesz and Aggozzotti, Rogers and Beritoff have shown the capacity of the decerebrate pigeon to learn, Beritoff (1) finding that it forms a simple conditioned reflex as rapidly as a normal animal. No completely decerebrate mammal has been subjected to systematic training for a long enough period to test the limits of its capacity for learning. Figure 1 shows the curve of improvement in an eight–cul-de-sac maze for four rats with an average destruction of 70.5 per cent of the cerebral cortex. The maze was not learned, but marked improvement occurred and the limit was not reached in the period of training. From such cases we may hazard a guess that, if shock could be avoided, the decerebrate mammal might acquire as complex habits as do fish or amphibia and that the learning capacity of lower centers has not been modified by the acquisition of the cerebral cortex.

The contribution of the different parts of the cortex to the formation of the maze habit has been studied by Lashley (24). Groups of animals with destruction of each of the chief cytoarchitectural areas were trained in the maze. Since the areas differ in size it was necessary to correct for such differences. This was done by computing from the regression coefficient for errors on extent of lesion for the entire group the expected errors for the magnitude of each area. The results were the following.

Errors	Motor	Somaesthetic	Visual	Auditory	Primitive associational
Experimental	468	414	649	148	521
Computed	484	597	570	145	598

The departure of experimental from computed values here is small in proportion to the difference from normal controls which made only 47 errors under the same conditions. No difference in behavior in the maze could be detected between groups with different loci of lesion. The experiment indicates that the differentiated areas of the cortex all contribute in the same way to the formation of the maze habit and in proportion to their areas.

Quantitative data for other animals are not available. Jacobsen has shown that destruction of the frontal or parietal association areas does not retard the formation of simple habits in rhesus monkeys, although the animals may have trouble with complicated problem boxes. For the learning of other problems the data are difficult to evaluate. In the rat the habit of brightness discrimination and certain problem-box habits are formed at normal rate after destruction of the cerebral hemispheres up to at least 50 per cent, and in no case has a greater retardation in the formation of a habit resulted from injury in one place than in another.

In man there seems to be a more definite restriction of learning by destruction of local areas. Improvement in every sort of activity seems possible after lesion in any locus (unless perhaps the total destruction of a projection area is involved) if the patient is not senile and can be prop-

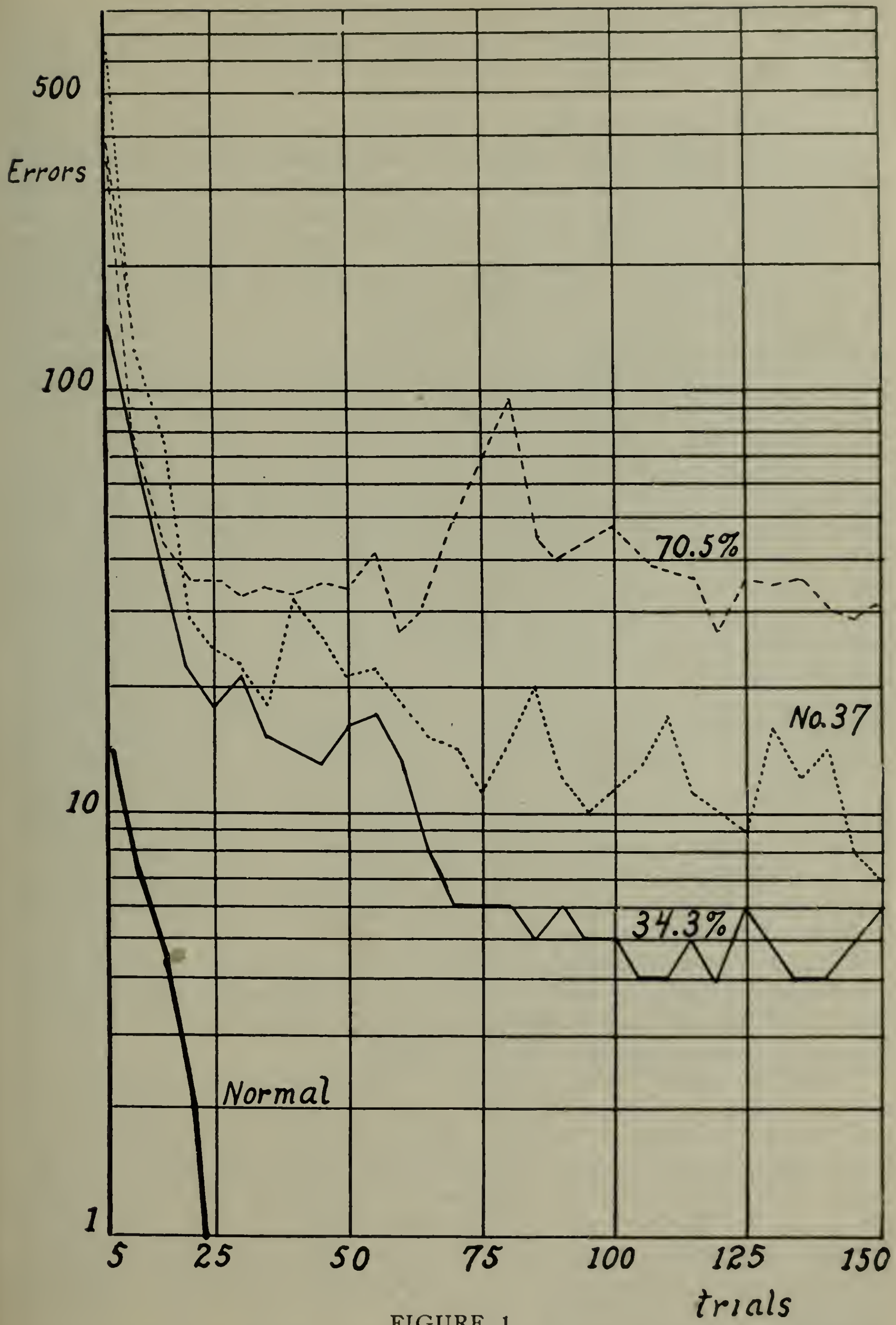


FIGURE 1

LEARNING CURVES FOR ANIMALS WITH EXTENSIVE DESTRUCTION OF THE CEREBRAL CORTEX COMPARED WITH CURVES FOR NORMAL ANIMALS

▬ Average of normal animals; —— average of four animals with approximately 34 per cent of the cortex destroyed; ---- average of four animals with about 70 per cent of the cortex destroyed; .... record of one case with more than 80-per-cent destruction. After Lashley (24).

erly motivated, but the rate of improvement in certain functions may be retarded. Thus the patient with severe aphasia seems to have much greater difficulty in forming verbal associations than in learning non-verbal material.

This does not prove, however, that the areas, destruction of which produced the aphasia, were particularly well adapted for the formation of verbal associations by virtue of any internal structural peculiarity or that other areas may not serve as well for the acquisition of verbal associations. The rate of learning in human adults is largely dependent upon pre-existing associations, as indicated by the importance of mediate associations in the learning of series of unrelated words. If all word associations are abolished by a cerebral lesion, slow learning is to be expected even though the vicariously functioning parts are structurally as well adapted for the learning as was the original speech area. Much of the difference between the results with the animals and man is probably referable to the use of acquired mnemonic aids by the latter.

Localization of the Engram. We have progressed a long way from the notion that individual ideas are stored in single cells of the brain. Ideas are recognized as composites involving the relation of many elements and dependent not upon one but upon many cells. Yet in a way this advance has only transferred the idea from the cell to the synapse, for the current theories of the dependence of habit upon specific neural arcs still localize the essential element of the habit in particular intercellular junctions. Are the engrams so definitely localized? Different experiments seem to give different answers to the question. In the rat the habit of brightness discrimination is abolished by destruction of the occipital lobes. Something essential to the performance of the habit is certainly located there. But it is not definitely localized within the general region, for smaller injuries within the area do not abolish the reaction but only weaken it in proportion to their extent. For the maze habit, on the other hand, there is no evidence for localization. It is weakened after any injury to the hemispheres but survives any small injury and is abolished by any large one, irrespective of location.

For man a great variety of habitual activities are known to be dependent upon particular areas of the cerebrum, but the localization is rather gross. Very small injuries rarely produce symptoms; larger injuries lower the efficiency of many complex related functions.

Clearly no habit can be ascribed to conduction paths restricted to a narrowly limited cerebral area, but beyond this the data are capable of at least two interpretations. It may be that each habit is mediated by a number of equivalent and scattered conduction paths and that efficiency of performance is reduced but not abolished by destruction of some of these. Or it may be that performance is not dependent upon the conduction of impulses over definite paths, but is dependent upon some general property of the propagated disturbances, such as periodicity of discharge, irrespective of the paths followed through the cortex.

Critical experiments upon this point are not available. Lashley (24) has reported cases where the interruption of projection fibers to certain cortical areas resulted in less disturbance of behavior than the destruction of the cortical areas supplied by the fibers, which would indicate that the conduction path is of less importance than the area to which it conducts, but the experiment may be susceptible to other interpretations. This problem of specific paths will arise again in the discussion of the characteristics of coordination involved in habit. From the anatomical and physiological data we cannot conclude with any confidence that the alterations of nervous structure underlying habits are located in any particular part of the nervous system or of the reflex path.

Significance of Finer Structures. The general belief that the neuron alone of the structural elements of the nervous system participates in the coordination of activity is supported by a large mass of indirect physiological and anatomical evidence, yet is not established beyond question. Cajal has suggested that differential response may be determined by the movements of neuroglia cells, their processes serving as insulators when extended between the neurons. There is no evidence to support such an hypothesis, beyond the variable forms of the neuroglia, but the suggestion, coming from an eminent neurologist, illustrates how little we know of the medium in which the neurons lie and of its possible influences upon their excitability.

The surface membranes separating the neurons have been most stressed in theories of the mechanism of integration, since they form the only known break in continuity of fibers where diffuse conduction seems likely to occur. The fundamental importance of the synapse seems a logical conclusion, yet we must bear in mind when evaluating theories of learning that the properties of the synapse are still entirely hypothetical. If we deduce its properties from the facts of learning, we gain nothing by explaining learning in terms of these hypothetical properties.

The recent work on nerve conduction seems to limit the possible theories of integration and of learning. If the propagated disturbance in the nerve-fiber may pass equally well in either direction, is subject to the all-or-nothing law, and is capable of modification only at the junction between cells, then the whole theory of learning must be worked out in terms of such processes. It seems well established that volleys of nerve-impulses are set up by sensory stimulation, their number and rate of succession being determined by the intensity of the stimulus, and their spatial distribution by the number and position of the end-organs stimulated. Their further distribution is determined by the state of excitability of succeeding neurons in the chain, by the intensity of the impulses (a product of the rate of succession and condition of excitability of the conducting neurons) and by the resistance offered by postulated regions of decrement (a resultant of initial condition and contemporal excitation from other sources).

Actually these facts have little relevance to the problem of learning since they leave it an open question whether the alteration of learning takes place in the excitability of the cell, in its rate of recovery from the refractory phase, or in the conductivity of the intercellular membrane;

whether the phase relationships, the intensity of discharges, or the resistance to transmission is altered. Furthermore, we do not know whether the laws of conduction in nerve-trunks apply within the gray matter. It may be that they do, but it is also possible that diffuse electrical or chemical processes within the non-medullated regions are equally important with the propagated disturbances in determining integration.

The data on the relation of nervous structure to learning offer little of constructive value. It is probable that simple associations may be formed in any part of the nervous system, so long as there is continuity of tissue between the afferent and efferent tracts associated. The cerebral hemispheres are somehow concerned in the formation of the more complex associations, but the nature of their contribution to learning is still a complete mystery. Special parts may be essential to the retention of a habit once formed, yet learning of that habit is not impeded by their absence. Learning of other habits may be retarded in the absence of parts of the cortex yet with no evidence of any localization of the habit mechanism in those parts when it has been formed. The cerebral hemispheres present no peculiarities of finer structure either in the character of the cells or in their connections, which can be pointed out as of especial importance for the learning function, and the simple mass of tissue is the only feature which thus far has been correlated with rate and complexity of learning.

General Characteristics of the Organization of Habits

We must now analyze more in detail the character of learned activities. It is customary to speak of them as reactions to stimuli, without analysis either of the nature of the reaction or of the nature of the stimulus. Thus an animal is said to learn a positive reaction to the door of a problem box; a familiar face is a stimulus to the pronunciation of a name. As brief descriptions of behavior such statements are permissible, but for the formulation of the learning problem they are very misleading. Just what is the stimulus in terms of the sensory fibers excited; what is a positive reaction or the pronunciation of a name in terms of muscular contraction? If we are to understand the mechanism of the habit, we must deal with it in these terms, for the integration is evidently between sense-organs and effectors and not between doors and directions of movement.

The Nature of the Stimulus. The theory of the specificity of conduction paths in habit demands that the reaction learned be given only to the excitation of the same receptor cells as were stimulated during training, or at least a significant number of them, these receptor cells being connected with the effectors by a system throughout which there is a point for point correspondence. These conditions are rarely if ever met in any training experiment. On the contrary, it is clear in most cases that the response is, within wide limits, independent of the particular sensory cells stimulated.

Figure 2 illustrates this in an experiment with rats. The animals, incidentally lacking the visual cortex, were trained to jump to a platform presenting a black surface and a white edge, seen against the background of the room. The appearance of the platform was then changed as indicated in the succeeding figures, by placing cardboard screens before it, or by substituting a large white screen on which paper figures were pasted. The animals jumped to any horizontal rectangle, whether black or white, but would not jump to a vertical rectangle or a plain figure. It seems clear that the stimulus here cannot be described as the excitation of such and such retinal cells. The elements common to the various situations are the proportions of the object seen against a varying background.

The writer previously reported the transfer of the habit of brightness discrimination from one eye to the other which had not been used during training. In this experiment the same sensory cells were certainly not stimulated in the formation and in the performance of the habit. Where is the common conduction path for the two sensory systems established by training? By one theory corresponding points of the two retinas are supposed to be projected to corresponding positions in the superimposed layers of the stripe of Baillarger. Here one set of retinal fibers might, by

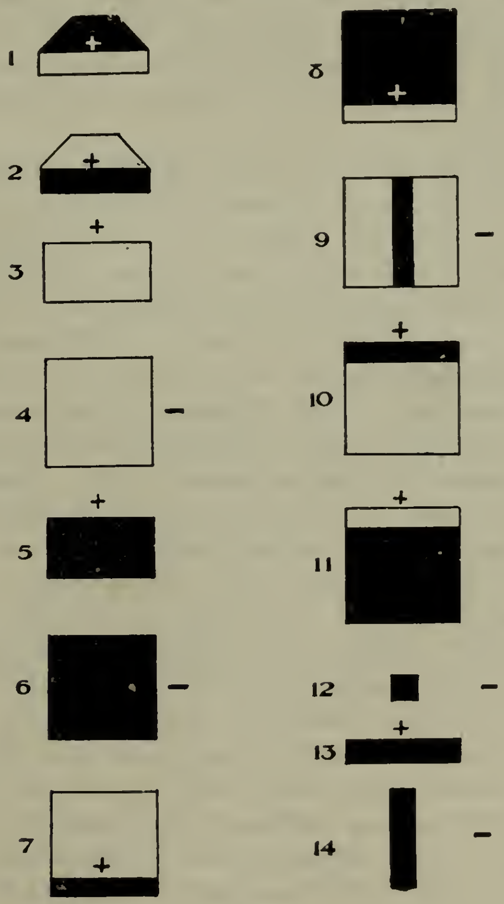


FIGURE 2

Equivalent Stimuli for a Habit of Jumping in the Rat

Animals were trained to jump to a platform having the appearance of 1, seen against the background of the room. The following figures were then presented in succession. The + sign indicates the position to which the animals jumped promptly; the —, failure to jump in five minutes.

drainage, condition the corresponding fibers from the other eye. But the experiment also showed that the transfer took place readily in an animal without a visual cortex. Of course, as long as we know nothing of the actual central nervous processes involved, the common neurons may be postulated in any remaining nervous tissue, but this is only an appeal to ignorance to save a weakened theory. If these were isolated instances, it would be worth while to try to explain them away, but such conditions are the rule, not the exception. We may state as a general principle that the stimulus to any reaction above the level of a spinal reflex involves, not the excitation of certain definite sensory cells, but the excitation of any cells of a system in certain ratios, and that the response may be given to the ratio, even though the particular cells involved have not previously been excited in the same way during the formation of the habitual reaction.

The performance of the habit cannot be ascribed to the functioning of stereotyped paths from sensory to effector cells. This characteristic of the stimulus has been emphasized by Driesch, Bergson, Wertheimer, Koffka, Goldstein, and many others, and, although we need not accept the anti-physiological conclusions which have sometimes been based upon it, we must recognize the problem as a serious one for psychological and neurological theory.

The Nature of the Reaction. In the simplest habits, such as the conditioned reflexes, the activities associated with the conditioned stimulus are those which were originally called out by the primary stimulus. This is essentially true of all habitual responses, but only if the word "act" is used in a broader sense than the contraction of a definite pattern of muscles. The rat learns to run the maze. Is the maze habit a chaining together of those movements which were used to traverse the maze during learning, or is it something else? A study of rats with severe motor incoordination from cerebellar lesions (Lashley and McCarthy) shows that, once the maze is learned, it may be traversed by a method of progression which involves no patterns of muscular movement that can be recognized as identical with those utilized in learning.

In a study of the effects of destruction of the motor cortex, Lashley has reported the immediate transfer of a complex pattern of movements from one arm to the other, which was paralyzed throughout the course of training, concluding that conduction paths used in this performance could not have been used during training. Herrick (15) has criticized this conclusion on the ground that since the impulses controlling habitual movements do not all pass through the motor cortex, the necessary impulses may have reached the motor paths of the paralyzed arm, even when the arm was inactive. But in other cases it was found that the head or leg might be used directly for the solution of the problem. This seems to demand either the assumption that during training specific conduction paths are laid down to every effector, whether it is used or not, or that the habit is not dependent upon specific paths.

As in the case of our conclusions concerning the nature of the adequate stimulus to habit, this type of behavior is the rule rather than the exception in habitual behavior. Similar instances are so common in human behavior that we need not cite specific instances. The facts are those which are held to characterize behavior as purposive rather than mechanical and, although we need not admit the finalistic implications of "purpose," we must recognize that a definite type of neurological problem is involved and that the behavior does not have the commonly admitted properties of the reflex.

Attempts have been made to meet this sort of problem by explanations based upon "symbolism." A stimulus is said to be associated with a verbal or other symbol; another stimulus arouses the same verbal reaction, and the two stimuli are therefore identified. Or the verbal reaction serves as a common stimulus to the various motor responses which may be used alternatively in the performance of the reaction.

If it could be demonstrated that similarity among stimuli is never recognized unless each has previously been associated with the same symbolic response, this explanation would be sufficient, but there is clear evidence that the use of symbols depends upon the recognition of similarity, and not the reverse, so that the identification of the symbol with the object involves exactly the problem which it was proposed to solve.

The appeal to symbolism is a variation of the doctrine of common elements which has been used widely to explain the transfer of training. It is held that the stimuli may be diverse but certain elements of each activate the same afferent paths, exciting identical nerve-cells and so eliciting the same reactions. Or transfer of training takes place because certain efferent nerve-cells supply two motor paths (the accepted theory of crossed education).

The conditions of visual stimulation which we presented on page 542f. seem absolutely to preclude any such common neurons, and it is equally difficult to find them in many cases of motor transfer (for example, the rhythm of two against three, once acquired, may be directly imposed upon any independent pair of muscle groups). The common elements in transfer are not common neurons, but common ratios of excitation in different neuronic systems.

The Nature of Central Integration Deduced from the Defects Following Injuries to the Nervous System

Studies of the symptoms of destruction of tissue within the central nervous system point to two seemingly conflicting principles of organization. There is evidence for a high degree of specialization for some functions corresponding to the conceptions of exact cerebral localization, and also evidence for an entire lack of specificity in the contributions of different structures to other functions. The evidence for exact localization arises chiefly from injuries within the anatomically defined sensory and motor projection areas of the cortex; that for diffusion of function, from the injuries to the association areas and from the more careful studies of symptomatology in all lesions.

The evidence for exact localization is not conclusive, even for the projection areas. Small lesions within the motor areas produce only temporary paralyses (Leyton and Sherrington), and the variations in function from time to time suggest that some of the apparent specificity of function of the motor cortex is a matter of temporary physiological organization rather than of permanent anatomical structure.

The doctrine of the "cortical retina" first proposed by Munk has led to the most extreme views of specificity of function of the cortical cells, up to the claim that there is a cell-to-cell correspondence between rods and cones and the ganglion cells of the area striata. Observations upon scotomas following occipital lesions show some correspondence (16), but as Poppelreuter (34) has pointed out, the variety in form exhibited by scotomas is far less than would be expected from the variety of lesions and may be interpreted as a centripetal or centrifugal irradiation of a disorganizing process within an undifferentiated area, as well as by the assumption of a cortical retina.

Data reported above (page 538) show that in lower animals habits are either not localized, as in the case of the maze, or only very vaguely in a general area, as in brightness discrimination. It appeared also that the capacity to learn and to retain habits formed after cerebral lesion may be a function of the total quantity of functional tissue.

For higher forms, including man, there are no comparable data. There are no controlled studies of the rate of learning in men with cerebral lesions, nor are there any attempts to take into consideration the extent of injury in evaluating symptoms, beyond vague statements that symptoms are slight or absent if the lesions are small. Inspection of the data on motor aphasia reviewed by von Monakow suggests that there is a fairly close relationship between the amount of destruction and the time required for recovery, but the data on rate of improvement are inadequate.

We may approach the problem of specificity of function in man in another way by inquiring into the general nature of the defects following injuries to the cerebrum. There seem to be three diverse types of disorder: simple sensory and motor defects, disturbances of memory, and disorders of the organization of functions. These distinctions are more apparent than real, but they serve to illustrate certain problems of neural function.

The purely sensory and motor disturbances have, as pointed out above, the closest association with structure. Franz (11) has shown that even these defects are not absolute, but are capable of a considerable degree of recovery and fluctuate in severity from time to time. There is some evidence that the motor defects are due to a weakening of some facilitating function rather than to loss of any definite coordinating mechanism. There may be, for example, a partial restitution of function during emotional excitement. The motor type of aphasia or anarthria is probably of this type. On the sensory side also, the suggestions of Poppelreuter concerning the nature of scotomas, the observations of Gelb and Goldstein (12) upon pseudofovea and the completion of figures in scotomatous areas are an argument against strict localization and in favor of a dynamic interpretation of the symptoms.

The notions of strict localization and limited defects in aphasia, apraxia, and visual agnosia were in large measure the result of inadequate analysis of the defects. The earlier observers noted the prominent symptoms, such as defects of speech, but had no means of testing other functions involved. They overlooked the milder degrees of dementia accompanying aphasia or ascribed them to involvement of other than the speech areas and made few efforts toward empirical analysis of the disorders of language and thinking.

The classification of aphasia is still a matter of controversy but many of the cases reported suggest two primary types of disturbances. The amnesic types present a picture of difficulty in the recall of verbal material, a difficulty which is more severe but not otherwise different from lapses of memory in normal individuals. In many cases recall is possible under favorable conditions. Speech occurs in excitement, or the proper word can be recalled with effort or by the use of mnemonic aids. Isolated words cannot be recalled, but come readily in certain contexts, or round-about expressions for the same thought are found. In general, there is not a selective loss of some words with retention of others but rather an increase in the threshold for recall of all verbal material. The exceptions will usually be found to be words having a considerable emotional reinforcement—the "emotional language" of Hughlings Jackson.

Such a condition is not to be accounted for on the assumption that specific associations have been interrupted. The lesion affects uniformly a large range of activities and reduces the efficiency of all without abolishing any. The areas concerned in speech seem to behave as does the whole cortex of the rat in the performance of the maze habit; injury reduces its general efficiency without selective effects upon the individual units of behavior.

Other aphasic symptoms imply a disruption of the normal mechanisms of organization rather than an amnesia. In extreme cases the speech is an unintelligible jargon; in milder cases grammatical form is lost, words are mispronounced or distorted, and the rhythm and orderliness of speech is disturbed. Such are the verbal and syntactical types of Head and the agrammatism of Pick.

The problem of the ordering of sequences of behavior, as in grammatical speech, is the most baffling of all neurological problems. It does not help to say that the order is a matter of habit for, although the general arrangements typical of any language are certainly acquired, the specific reactions employing any given word are not. If we learn a new word, such as the name of an object, that word will later be used in its proper relation to other parts of speech, although the context is one in which it has never before been heard. It is not linked to the other words of the sentence by habit, except in so far as it belongs to some general category of parts of speech.

To form any notion of the neurology of grammar we must assume that the entire mechanism involved in the verbalization of a proposition or idea is thrown into a state of partial excitation before actual speech, internal or overt, occurs; that in the preliminary activation the relations implicit in the idea are organized; and that this organization is impressed upon the verbal mechanism as from without. The problem is not that of the activation of successive links in a chain of associated reactions, but the determination of the order in a conglomerate of processes in which there are not previously formed associative bonds.

Jackson formulated this notion in his recognition of the importance of the proposition and Head has emphasized it in distinguishing the syntactical and semantic types of aphasia. The same sort of problem appears at every level of neural organization. Some of the difficulties in reading after occipital lesion have been ascribed to loss of spatial organization within the visual field. The patients see the letters or words but these no longer have a recognizable spatial position with respect to one another. Defects of insight and the intelligent manipulation of ideas evidently involve the same sort of confusion in organization and it seems possible to distinguish a continuous series of such disorders, ranging from serious intellectual defects to slight disorganization within sensory fields. In how far the defects are called general or specific seems largely a matter of interpretation. Head (14) considers that intellectual defect occurs in aphasia only in so far as the lack of language is a handicap. Bouman and Grünbaum (3), on the other hand, consider the aphasias as primarily a defect of intellect. "The patient is not able to hold in mind the concrete elements of a problem and at the same time manipulate them in thought." This difficulty is evident in all behavior of the aphasiac, and the difficulty of speech is only one expression of it.

Whatever the final conclusion as to the general cerebral localization of speech, it is clear that this localization does not extend to the logically differentiated elements of speech. These do not correspond to the physiological elements. The problem of cerebral function is characterized by the interplay of many organized systems in which equilibrium or dominance of excitation and the relations among the parts, rhythms and timing of activities, patterns of excitation, rather than of performed association paths, express the significant facts.

Another group of phenomena important for an understanding of cerebral mechanisms is indicated by the relative fragility of functions. In the visual field, for example, the discrimination of colors is more frequently disturbed than any other function and by slighter injuries. When recovery occurs from an initial severe hemianopsia, the visual functions reappear in a rather definite order. Poppelreuter lists sensitivity to light, recognition of differences in illumination in different parts of the visual field, recognition of the direction of single lines without identification of figures, recognition of simple geometrical patterns, vision for complex forms with confusion of meaning, and normal vision as stages in the recovery of pattern vision.

Attempts to localize these different functions in different areas have not been supported by anatomical findings and the stages of recovery or the relative ease of abolition of the functions can be interpreted only as an indication that the same area may function at different levels of complexity. The problem here is like that presented by cerebral function in maze learning, where a limit to the complexity of the habits which may be formed is set by the extent of the injury. With increase in mass and physiological development the complexity of the situations which can be integrated increases. We have as yet no clue to the nature of these "levels" of integration, except the evident importance of the quantity of tissue.

The symptoms of lesions in man emphasize the dynamic rather than the static in neural organization. There is a simplification of behavior, a reduction in the number of elements which can be dealt with or integrated in a single reaction. This simplification may be so general as to involve almost every aspect of behavior, constituting a general dementia, it may be limited to some one field of activity such as the verbal formulation of propositions or the synthesis of visual impressions, but in every case it affects the majority of reactions of the given type. It is not the specific reaction but the mode of synthesizing activity which is affected by cerebral lesion.

Along with this there may go an apparent weakening of the reaction mechanism. Reactions are not absolutely abolished but only rendered difficult, so that some extraneous facilitation is required before they can be elicited. Here also, not specific reactions but groups of reactions are affected and the grouping is determined by such relations as object–name. It seems as though the unit of cerebral organization can be expressed only in terms of relationships among activities, rather than in terms of individual reactions.

Peripheral versus Central Processes in Habit. The last decade has witnessed a bitter controversy concerning the nature of reproductive memory. Until the development of objective psychology the doctrine that memory involves the rearousal of central nervous processes reproducing more or less completely the original sensory experience was generally accepted. No clear neurological theory concerning their production was formulated beyond the general notion that they were produced by the action of centers which were thrown into activity by associative processes.

Watson (40) attacked the theory of imagery on the ground that memory-images are not objectively demonstrable, that introspective studies purporting to deal with them become involved in a futile mentalism which has no explanatory value for behavior, and that the phenomena of memory and thought may be dealt with more adequately on the assumption that recall is the reinstatement of those bodily movements which were first elicited by the stimuli recalled. Thus all memory was reduced to terms of motor habit, and thought to chains of motor reactions in which muscular contradictions serve as stimuli to further activity. Dunlap's theory of homeodetic arcs and Washburn's of "tentative movements" have much in common with Watson's doctrine of "implicit movement," but the former admits a central short-circuiting and the latter ascribes the image to the blocking of the motor response.

Recent attempts to establish the importance of implicit movements or movement systems (Washburn) in thinking have given negative results, while the evidence for centrally maintained coordinating processes has accumulated. Lashley and Ball (25) found that the performance of the maze habit was not disturbed by the destruction of the somaesthetic paths in the cervical cord and concluded that the maze habit cannot be interpreted as a chain of somaesthetic-motor reactions. Attempts to demonstrate implicit movements in the verbal mechanisms during thinking have shown movements, it is true, but they are irregular, by no means always present, and do not correspond in rhythm or pattern to those of overt speech (37). They can be interpreted only as chance irradiation to the verbal mechanisms from centrally maintained processes. The writer has obtained similar results in a study of eye movements during thinking of geometrical forms and relations.

The argument that the interference with thinking by maintained postures of the articulatory organs, as in the difficulty of thinking "bubble" with the mouth widely opened, is an evidence for peripheral interference is not valid, for a similar interference can be demonstrated between different motor organs where the blocking can be only central. For example, it is difficult to tap a rhythm of three with one hand against two with the other, although these processes can be dissociated by practice. Similarly, Pintner showed that verbal thinking could be dissociated from the interference of articulatory activity. Reports of the effects of operative removal of the larynx or other articulatory mechanisms do not indicate any resulting disturbance of thinking.

On the other hand, something like objective evidence in support of imagery is accumulating. Visual after-images have never been denied by the opponents of the memory-image, but have been discussed as genuine phenomena and ascribed to residual excitations within the peripheral mechanisms. Recent studies of eidetic imagery reveal objectively a very close parallelism between the eidetic image and the visual after-image. Naïve subjects, who could not possibly be familiar with the laws of color mixture, report phenomena of color contrast, changes in the apparent size of the image when projected to different distances, and the like, which can be explained only on the assumption that there is aroused in them by association some process which corresponds very closely to the excitation originally aroused by peripheral stimulation (21).

Since the motor theory of imagery assumes that recall is a reinstatement of the reaction originally given to the stimulus recalled, the image should contain no elements which were not reacted to or noted in the original situation. Woodworth has held that this is true of the image, but some of the descriptions of the content of images given by eidetic subjects include a wealth of detail which almost certainly was not noted by specific reactions during the period of exposure of the stimulus.

Although perhaps not conclusive, this evidence seems to favor the older doctrine of imagery and to throw us back upon the concept of activity maintained within the central nervous system for an understanding of serial habits and the mechanisms of thinking.

The Phenomena of Learning

The foregoing discussion serves to define some of the characteristics of the learned reaction and the properties of the mechanism by which the performance is brought about. Not all experiences come to be associated and, since we cannot observe directly the nervous changes, we must fall back upon analysis of the external conditions which favor association and try to infer from them the nature of the central nervous processes which occur during practice.

Psychological studies have given a series of "laws of learning" which define conditions under which learning occurs and which have served as the points of departure for theories as to the nature of the process. The laws of contiguity, exercise, effect, and completeness of response are the ones most stressed in recent literature. In addition, primacy, recency, vividness, intensity, and the older laws of association, such as similarity, have been cited by various writers as significant for the theory of learning. Since apparent exceptions to each of these laws are numerous, we must examine them briefly before admitting that they form an adequate basis for learning theory.

The *law of contiguity* seems the most clearly established of these generalizations. In order that association occur, the associated experiences must fall within a certain time interval. For conditioned reflexes in lower animals this interval is a matter of seconds or minutes; for man, perhaps through mediate associations, it may be much prolonged. Exceptions may be the seeming facts of association by similarity, as when a visual pattern suggests an auditory rhythm. Since not all experiences which occur simultaneously are associated, the law clearly defines only one of the conditions necessary for learning.

The *law of exercise* has been more emphasized than any other in recent theories. Repetition of experiences in contiguity leads to their firmer association. Repetition of a function results in its improvement in speed or accuracy, according to the law of diminishing returns. These statements express common facts. But their universality and fundamental importance for learning theory is questionable. Most of the learning of human adults does not involve repetition and frequent instances of animals' learning in a single trial are available. Moreover, the functions which are improved by repetition are invariably complex, involving the formation of numerous associations, many of which may be seen to be formed in a single trial. The learning curve may be explained as well on the assumption that single associations are formed in one trial and do not improve with practice, as by the theory that they improve gradually. Thorndike (36) has so dealt with the learning curve in discussion of changes in the rate of improvement, although he has elsewhere adhered to the doctrine of gradual improvement in the strength of bonds. Cason (5) has pointed out that improvement in the function usually involves not an improvement in the specific acts but a change in the character of the acts; that this is the establishment of new associations, not the improvement of old. Thus for the greater number of habits it is questionable whether improvement involves a strengthening of associations.

The facts which seem best to support the law of exercise are those revealed by the savings method and nonsense material. Repetition of the series seems to strengthen the associations, since, with increased practice, less and less is lost by equal periods of disuse. This would be a more impressive argument if nonsense syllables were truly nonsensical, but actually their acquisition is helped by various mnemonic aids, and it is quite possible that overlearning multiplies the number of such associative bonds rather than strengthens those first formed.

The improvement during conditioned-reflex experiments, involving apparently the association of a single stimulus and response, might be interpreted as evidence for improvement in conduction over a single path, but a considerable part of the improvement, which is measured by a greater regularity in the appearance of the reflex, is obviously due to a blocking of specific inhibitions, and it is not clearly demonstrated that any strengthening of bonds occurs even here.

Data on "fatigue" or experimental extinction of the conditioned reflex, and the phenomenon of "going stale" from overpractice show that repetition may weaken rather than strengthen associations. Pavlov (29) has contended that this is due to the development of inhibition, on the ground that it cannot be a fatigue of the conditioned-reflex path. The term inhibition is meaningless in this connection, since we are not dealing with the suppression of one activity by another, which is all that inhibition means today, but with the suppression of an activity by simple repetition. The phenomenon seems related to that of adaptation in lower organisms, to which there is still less reason to apply the notion of inhibition. Experimental extinction stands today unexplained and in opposition to the law of exercise.

It is not our purpose to minimize the importance of exercise for learning, but only to point out that the facts of improvement with practice are susceptible of various interpretations and do not justify the usual inferences drawn from them concerning the nature of the nervous processes involved. Associative connections are strengthened by practice but it does not follow that these connections are nerve tracts in which resistance is lowered by repeated passage of impulses. On the contrary, the changes in the character of the responses with practice indicate the continuous formation of new associative bonds, and there seems to be no instance of learning on record to which this explanation of the law of exercise could not apply.

The *law of effect* has been variously expressed. Beginning as the notion that pleasant or unpleasant results of an act determine its repetition, it has evolved through successive stages to the theory of the fixation of habit through the influence of the "consummatory reaction." The pleasure-pain theory has been criticized on the grounds that it is mentalistic, that it begs the question by assuming that the consequences for every case of learning have been pleasant, that it fails to account for the acquistion of likes and dislikes, that it fails to account for selective association in such cases as maze learning, and that it is not true for reproductive memory. in which unpleasant experiences are recalled as readily as pleasant.

In many of the studies of learning in animals it is customary to use incentives such as hunger, punishment, and the like, and in most cases the acts learned are those which result in removal of the incentive. The predominance of this type of situation in experimental studies has led to a great emphasis upon the importance of the "drive" and the "consummatory reaction." It has been assumed that the hunger or other organic need serves to increase the general tension of the organism and that the cessation of the tension, following the consummatory reaction, with its satisfaction of the need, somehow fixes the habit.

Two major theories of the mechanism of the drive have been implied in recent writings. The one takes its departure from the hunger mechanism and assumes that a continued state of visceral tension supplies a mass of sensory impulses which, by irradiation or specific facilitation, increase the excitability or responsiveness of the nervous mechanisms involved in exploratory behavior. This notion of organic hunger is probably correct, but it does not provide a comprehensible mechanism for the fixation of habits. The action of glands of internal secretion has been suggested as a second factor in the drive on the assumption that the heightened activity during emotional disturbance is referable to the increased adrenalin output and that other endocrine organs may work in the same way. We have carried out a number of experiments with adrenalectomized and castrated animals and other tests in which attempts were made to swamp the normal variations in secretions of the pituitary, adrenal, and sex glands by subcutaneous injection of extracts. All gave negative results on learning.

A good bit of evidence, such as the studies of relation of strength of incentive to rate of learning, shows that the drives play an important part in the formation of habits, but it is not clear whether this is by heightening the general activity of the animal and so increasing his chances of exposure to the proper conditions for learning, by increasing excitability so that the effects of the associated stimuli are more intense, or by some specific action whereby the associations are fixed. Certainly what is known of the mechanisms of the drive provides no comprehensible mechanism for fixation. Moreover, in the greater part of human learning, particularly in incidental memory and recognition, there is no evidence for any factors comparable to the drive.

The assumption has been based upon the successive formation of conditioned reflexes that the drive may be transferred from one reaction to another and that the primitive drives may thus play a part even where they are not apparent. The latter conclusion is, however, purely speculative and is really an attempt to explain the more common phenomena of learning in terms of the less.

The law of effect, in its various forms, confuses learning with performance. The animal learns to go where food is obtained in the maze and when hungry goes there. Under other conditions he may show an equal familiarity with other paths in the maze. The evidence suggests that the drive is one of the associated elements in the maze habit rather than the agent responsible for the association.

A view of the learning problem which is allied to the emphasis on the law of effect is that stressing the unity in the act of learning. Peterson (31) has pointed out that maze learning and similar activities cannot be understood in terms of independent *acts*, but must be looked upon as a unitary function in which all behavior is fused into a harmonious whole. The perfected habit is more complete, in that it involves fewer conflicts of impulses than the initial random activity, and the learning is the selection of these more complete responses. Poppelreuter (33) denies that simple association ever occurs. Learning results from the "coexistence of experiences in one total experience whose physical correlate then constitutes the engram." These statements foreshadow the doctrine of "closure" of the configurationists. Koffka (22) has developed this notion in some detail. In solving a problem the organism is working toward a definite end or goal which is unknown but nevertheless effective in producing activity. Until this goal is reached, the activity is incomplete or "unclosed." With the attainment of the goal the series of activities is completed or closed; that is, it assumes a configuration or equal distribution of tensions which is normal for the organism, as the spherical form is normal for a bubble. With closure, retention is somehow assured. Koffka's statement of the law of closure parallels closely the law of effect. The unclosed situation corresponds to the drive to activity, the closure to the consummatory reaction. At present the closed system is not defined except in terms of analogy with physical systems in equilibrium. Underlying the theory of closure is the notion that the nervous system is so constituted as to limit and prescribe the relations into which excitatory processes must fall, as the shape of a surface determines the distribution of a static charge upon it. Description of insight is description of these necessary configurations, although in terms of stimulus relations or sensations. We cannot find that any clear positive statement of the neural processes has been given. Wertheimer speaks of direct relations between centers of excitation, independent of the specific intercellular connections but does not elaborate the suggestion. The recent emphasis upon the concept of configuration has been valuable in bringing out the ubiquity of the problem of reaction to relations in psychology and the inadequacy of the simple mechanical hypotheses to meet the problems, but it has provided no adequate substitute for these explanations. As applied to learning, there is no present criterion of closure of a system of activity except the fact of learning so that, if the hypothesis does not beg the question, it at least is inapplicable in advance to any particular learning problem.

The *laws of primacy and recency* have not usually been considered as fundamental conditions for learning. The theory of backward association of Hachet-Souplet assumed that when a series of acts leading to food is learned, the last act of the series is first associated with the food, and in the following practice the preceding acts are associated in order. Studies of the order of elimination of errors in the maze have not supported this view of backward association. It seems probable that primacy and recency are effective only when they increase the intensity or stimulating value of the situations.

The *laws of intensity and vividness* seem to apply more widely than any of the other generalizations concerning conditions favoring learning, with the exception of the law of contiguity, but this is perhaps only because we know so little about them. Within limits, the rate of formation of the conditioned reflex varies with the intensity of the associated stimulus. The rate of learning of a discrimination problem varies with the intensity or stimulating value of the stimuli, but here at once we meet with difficulties, for although factors in brightness discrimination can be stated in terms of the energy of the stimulus, it is evidently not the energy of the stimuli which makes one pattern more easily learned than another. The effectiveness of a situation as a stimulus can rarely be stated in terms of objective intensity and only such vague notions as the value for attention can be used to define the subjective intensity or vividness. Yet it seems clear that the effectiveness of situations as stimuli is of prime importance for their association. A few exceptions complicate the problem, as the formation of associations with subthreshold stimuli. There are not data upon which to base a neurological theory of stimulating value. On the whole, the problem seems more closely allied to that of logical memory than to the mere intensity of excitation.

The importance of the *logical relations* of the material to be learned is shown by a large number of studies. The use and value of mnemonic aids is sufficient evidence of this. There is, however, little evidence as to how the ordering of the material to be memorized favors learning. The most commonly expressed theory implies that logical learning requires fewer new associations than rote learning of an equivalent amount of material. Meumann has expressed this in the statement that "learning is not a mere matter of the number of elements but of the number of independent memorial units." It seems doubtful whether this interpretation can be substantiated. The following verbatim statement from a subject with an unusually good memory for numbers shows the multiplicity of relations which are recognized and retained in the memorizing of a simple number. The subject was asked to remember the number 8264 and replied, "That's easy because my telephone number is 8238. The first two numbers are the same: multiply by 2 and divide by 2—that is, the last numbers—gives 6 and 4: then 8264 goes down in steps of two, except that the 2 is transposed to the second place." These are, of course, familiar relations for the subject but they are in an unfamiliar order. Where to multiply or divide is dependent on the new number, and the order of these procedures must be remembered. The assumption that few associative bonds are required is wholly gratuitous. Where large groups of numbers are remembered as familiar units, there is doubtless a reduction in the number of new associative connections, but such are no more instances of logical memory than is the learning of nonsense words instead of their constituent letters. The logical ordering and memorization of material seems always to involve the construction of some sort of relational framework and the association of the elements to be remembered with parts of the logical frame: a process which would seem to involve more rather than fewer associative bonds.

In the discussion of the symptoms of cerebral lesion in man we have seen that the functional units must be regarded as modes of organization rather than single reactions, methods of relating rather than specific associations. There is a resemblance between this and the phenomena of logical memory, as though the material to be memorized were fitted into some one of the modes of organization, and as though retention were better if the material fits into some general schema of facilitation such as we have suggested to be functional in determining grammatical form.

It is difficult to distinguish logical from rote memory. It is not impossible that even the supposedly most mechanical learning, such as the rat's acquisition of the maze habit, involves something akin to logical memory. There is certainly evidence of a generalization of directions in the maze as an important factor in the formation of the habit. There seems to be as much justification for Poppelreuter's conclusion that all learning is of the type of logical memory, as for the opposed view that all learning is the formation of fewer or more conditioned reflexes.

Theories of the Mechanism of Learning

Attempts to formulate a physiological account of the learning process have been concerned either with descriptions of the supposed structural changes underlying new modes of behavior or with speculations concerning the processes by which such structural changes are brought about. Occasionally the two aspects of the problem have been included in one theory. The speculations concerning the nature of the memory trace or engram may be divided roughly into five groups.

1) *Growth of New Processes Connecting the Neurons.*[1] Earlier views of the dendrites as amoeboid processes, capable of extension and contraction in the establishment of temporary anatomical connections, led to the supposition that learning might consist of alterations in the contractility of the cells (Cajal, Hellwig). Sherrington's demonstration that reflex latent time does not involve a preparatory setting of the path, with more recent studies of finer cellular structures has led to the abandonment of these theories. More recently Kappers (20) has suggested that learning involves the establishment of new relatively permanent connections through the growth of axons and dendrites under the influence of bio-electric currents. His general theory of neurobiotaxis was sketched on page 527.

Its application to the problem of learning is illustrated in Figure 3. $a - A$ and $c - C$ represent two conducting systems, and β a cell of a sensory tract, not associated with either but in close proximity. If

---

[1] A good review of the anatomical theories is given by Matthaei (27).

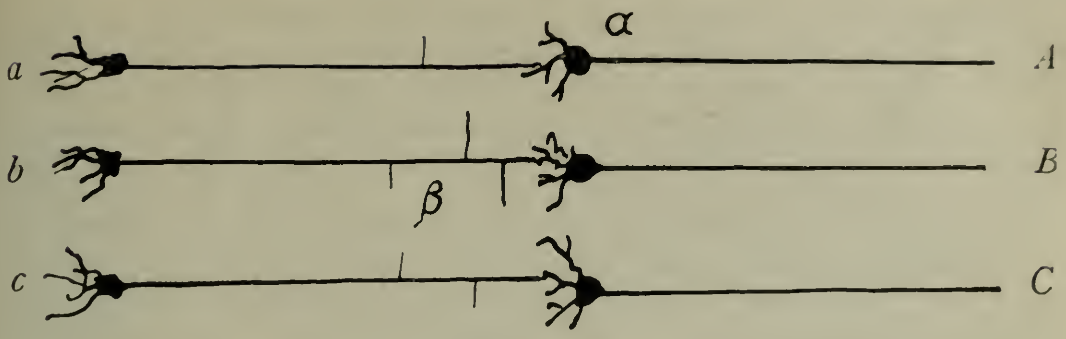


FIGURE 3
Diagram to Illustrate Kappers' Theory of Association
For explanation see text.

a is excited, cell *α* is thrown into a condition of heightened "electrolytic dissociation" so that the surrounding field is ionized. If now *β* is stimulated, growth of its axon and collaterals is induced. The growth occurs most strongly in the ionized field of *α*, rather than toward any of the cells of *c* – *C*. The approach of a collateral from *β* stimulates development of dendrites from *α* until contact is finally made.

Certain inadequacies of the theory are apparent. It accounts for selective association between two paths, but does not account for the direction of association. The rate of amoeboid movement and cell growth seems too slow to account for the speed with which associations may be formed.

2) *Increase in Conducting Substance as a Result of Exercise.* Verworn argued, from the growth of exercised muscles and the atrophy of the cells of the visual cortex in blinded animals, that the nerve-cells increase in size under the influence of stimulation, and that with increase in size there is a corresponding increase in the intensity of the nerve-impulses. Others have stressed the growth of parts of the cell, the axon (Exner), the neurofibrillae (Ziegler). Alterations in the conducting substance without implied change in the morphology of the cells have been discussed by many writers. Robertson, from similarities between the experimental curves of Ebbinghaus and certain curves of rate of chemical action, has deduced that learning consists of the deposition within the cell of a product of an "autocatalytic monomolecular reaction," the diffusion of which through the cell membranes constitutes forgetting. Matthews and Crile have pointed out the similarity between hysteresis of linseed oil and the phenomena of memory, and have suggested that since linolinic acid occurs in the central nervous system it may be responsible for learning.

3) *Chemical Changes in the Regions of Contact between Cells.* The most widely cited theory of the engram is derived from Sherrington's theory of the synapse as a semipermeable membrane whose resistance to the passage of nerve-impulses is primarily responsible for the patterns of integration. From this it was but a step to the conception

of a change in the permeability of this membrane as the basis of memory. Other theories concerning this mechanism have been suggested. Goldscheider postulates specific molecular changes at the intercellular junctions, by which a specific chemical sensitization between cells is produced.

Cason (6) has enumerated many ways in which the cellular relations might be modified by the learning process. Reorienting the molecules of the synaptic membrane by electronic bombardment so as to alter its permeability, physical growth in accord with Kappers' theory, changes in the colloidal dispersion within the neuron, according to McDonald's theories of conduction, and changes in surface tension are the chief possibilities suggested. No evidence is given that all or any of these changes are actually involved in learning.

Johnson (19) has formulated a detailed hypothesis concerning the structure of the synaptic junction, of which only a brief and inadequate sketch can be included here. He assumes that ionized molecules are absorbed on the cell membranes, constituting a surface film in which the charged ions are oriented with respect to the membrane. The ordinary metabolic processes of the cells, involving a constant transfer of ions through the membranes, produce a continuous oscillatory change in the orientation of the molecules of the surface film. The rate of this oscillation is constant for any given set of conditions. In order that conduction from cell to cell should occur, the periods of oscillation of their films must be synchronous. The periods of oscillation are altered during the learning process, by a "forcing" of their frequencies resulting from the simultaneous stimulation of the membrane by the sensory impulses which later come to be associated. The hypothesis is chiefly of interest because of the modification of the theory of drainage which is introduced.

4) *Persistence of Excitatory Processes.* Ebbecke (10) has pointed out the resemblance between visual after-images, memory-images, and a wide range of other phenomena and, on the assumption that the simpler of these are due to an after-discharge, has attempted to account for all memory in terms of residual excitation in the cells.

5) *Diffuse Functional Correlations.* Loeb (26) followed Goltz to a large extent in rejecting cerebral localization of function and proposed a theory of learning in accord with this view. The theory was rather vaguely expressed but involved the assumptions that integration in the nervous system is determined by the periodicity of discharge of nerve-impulses rather than by specific connections of cells and that learning consists of a tuning of motor systems so that they become resonant to particular sensory frequencies. A similar conception of the mechanism of integration is proposed by Weiss (41) although he has not applied it to the problem of learning. This alteration in periodicity of activity may be ascribed either to individual cells making up distinct paths or to large masses of cells. Its locus has not been discussed by the writers.

This list of theories includes almost every conceivable change which could occur in a system made up of nerve-cells. There is little to choose between the different ones. They are not developed in detailed relation to the various phenomena of learning and retention and so are not readily capable of experimental test. No one more than another serves to unify and make more intelligible the diverse phenomena of learning. None provides any basis for prediction of the results in any new learning experiment.

Almost all of the theories take for granted the law of exercise, the formation of associations by the forced passage of nerve-impulses over the paths later associated, and the functioning of these paths in the performance of the learned activities. As we have seen in the foregoing sections, these are not established facts and indeed seem to be contradicted by a considerable amount of evidence. If it is true that a learned response may be carried out by the activation of conduction paths, either afferent or efferent, which were not employed during the process of learning, or if practice does not lower resistance to conduction in established paths, but only leads to the formation of different paths, the theories lose all their relevance.

The theories reviewed above have been concerned chiefly with attempts to describe the engram. Another group has attempted to account specifically for the initial passage of impulses from one tract to another, assuming that the law of exercise will explain the fixation. They start with the fact of simultaneous or successive excitation of two tracts which are in anatomical but not in functional connection. Drainage and irradiation of nerve-impulses are the two mechanisms most frequently cited.

Drainage. The theory of drainage was first definitely propounded by McDougall to account for the phenomena of inhibition. It involved the postulation of a particular kind of energy, "neurin," stored in or flowing through nerve-cells. When a neuron is stimulated, it not only conducts the energy received from the stimulating cell but also drains off the energy of other cells with which it has organic connections. The theory has been applied to the learning process by Meyer and by Adams. In brief, they assume that when two reflex paths are excited simultaneously the dominant final common path drains off energy from the other. The connections opened by the force of this drainage are later permeable to impulses coming from the previously drained path alone.

The theory of drainage is opposed by all of the more recent data upon the nature of the propagated disturbance in nerves. The phenomena which it was devised to explain can be more adequately treated in terms of interference of nerve-impulses. Johnson (19) has attempted to rehabilitate the theory in order to explain the induction of retinal fatigue in one eye by stimulation of the other. To the writer this seems only a special case of central elaboration in vision, presenting the same problems as binocular color mixture, filling in of the blind-spot, completion of forms in the hemianopic field, and the like. These phenomena cannot be dealt with in terms of drainage so that its postulation for the special case of fatigue and for learning seems unjustified.

Irradiation. The theory of irradiation as proposed by Goldscheider assumes that regions intermediate between two excited centers receive nervous impulses from both and are, therefore, subject to stronger excitation than non-intermediate zones. As a result of this intense excitation, the region undergoes changes of stable character which result in a more ready excitation of either center by the other. The more recent formulation of Pyle seems essentially the same as this.

Watson (40) has approached the problem in a somewhat different way. He assumes that the excitability of the final common path of the associated reaction is increased during stimulation. When the associated stimulus is applied, it excites sensory fibers, and weak impulses irradiating from these may stimulate the hyperexcitable final common path. This excitation results in a lowering of threshold for impulses coming later over the same sensory path.

The theory of Kappers (see page 556) also appeals to irradiation but assumes that the spreading electric current resulting from the nervous discharge stimulates the growth of processes from the cells.

Direct evidence for the theories of irradiation is lacking. The writer attempted a test of them by training animals under the influence of strychnine and caffeine. These drugs are supposed to have a similar physiological action, promoting irradiation in spinal nervous centers. Strychnine was found to facilitate learning, caffeine to retard it markedly, which proves only the inadequacy of our knowledge of pharmacology.

Learning by Trial and Error. Most of the recent attempts to develop a theory of learning have started from the problem of random activity and the selection of successful acts. The statement of the problem is usually some variant of the following: a stimulus (hunger, maze, or what not) initiates a series of activities which may be called *a-b-c-d-e-f-g-h,* and which follow each other in a natural order. Acts *b, d, f,* and *h* contribute to success in the problem. With practice, the others are dropped out, leaving the series *b-d-f-h* as a direct solution. No question of the association of *b* with *d* and so on is raised; it is taken for granted when *c* is omitted.

This we believe to be a wholly misleading statement of the problem of learning in the maze or problem box. The acts do not form an internally conditioned series from which the dropping of some will leave the remainder in association. On the contrary it is a chance sequence determined by the environmental stimuli which the animal encounters, an odor here, a contact there, calling out specific unrelated activities. The linking of the successful acts is not explained by the dropping of the unsuccessful.

The acts making up the final series are not those which originally intervened between the unsuccessful movements. Entering the true path from a section of the cul-de-sac in the maze is the same act as entering the section from the preceding part of the true path only when the act is defined in terms of its end result; the movements involved are never the same.

The first evidence of learning is a focalization of activity to which the elimination of errors seems purely subsidiary. In the maze the animal pushes forward in the direction of the food compartment; in the problem box his responses to the latch become prolonged. The effectiveness of certain stimuli is heightened, responsiveness to other stimuli drops out in consequence. The problem of learning by trial and error is thus not one of the mechanism of dropping of useless movements, but of the fixation of the effective behavior. The behavior shows a responsiveness to relationships and a unity which is difficult to express in terms of simple associations and for which none of the suggested principles of selection is a satisfactory explanation.

THE PROBLEM

The conclusions from this review of the problem of learning must be largely negative. In spite of the vast experimental literature which has grown up since the studies of Ebbinghaus and the advancement which has been made in the study of nervous structure, it is doubtful that we know anything more about the mechanism of learning than did Descartes when he described the opening of the pores in the nerves by the passage of animal spirits. His statement sounds remarkably like the reduction of synaptic resistance by the passage of the nerve-impulse.

The evidence from many lines of investigation opposes interpretation of learning as the formation of definite "conditioned-reflex arcs" through the cerebral hemispheres or through any other part of the central nervous system. It is doubtful whether the conception of limited conduction paths through the gray matter can be applied even in the case of spinal reflexes. The adequate stimulus to habitual reactions and even to instinctive acts is a ratio of intensities distributed in time or space and capable of calling out the response when applied to any sensory cells, within a wide range. Analysis of the central nervous mechanisms reveals likewise an independence of individual nerve-cells and a determination of response by masses of tissue and relative excitation of different parts. The same sort of independence of structural elements appears in a study of the motor responses. The phenomena can be described only in vague dynamic terms, and attempts to particularize either in terms of individual reactions or of anatomical units fail to express the most striking characteristics of the problem of organization. The unit of functional organization seems to be not the reflex arc of Sherrington or the mechanism of reciprocal innervation of T. G. Brown, but the mechanism, whatever be its nature, by which response to a ratio of intensities is brought about.

If our analysis of the problems of behavior and of neural mechanism is correct, it follows that the current theories of neural organization have started from false premises and offer no hope of a solution of the problems. Such a sweeping conclusion is still premature, but its possibility emphasizes the need for a more critical experimental analysis of the evidence upon which current theories are based.

A review of the literature on learning reveals many suggestive relationships between the various forms of modification of behavior. We have pointed out the similarity between the problems of morphogenesis and of integration in the developed nervous system; the relations of fatigue, acclimatization, experimental extinction of the conditioned reflex, and "going stale"; the problem of the organization of behavior in relation to differentiated structure and to undifferentiated mass of neural tissue; the interrelation of speed of learning, limits of training, and cerebral development; questions of the nature of the adequate stimulus and the seeming differences between learned reactions and the mere contraction of muscle groups; various instances where specific laws of learning seem not to apply. In none of these examples are the data adequate to give a clear insight into the relations involved, and until these relations are revealed there is no starting point for a general theory of learning.

In the present stage of psychological science, the sole value of hypotheses is to define alternative possibilities which can be tested by experiment. The slight value of current formulation of learning theory is attested by the small number of studies which attempt a critical evaluation either of the laws of learning or of the postulated nervous processes underlying them. The fact is that most of the laws and theories are couched in such form as neither to be susceptible of experimental verification nor to suggest any experimental procedures which will lead to a more profitable formulation. Descriptive studies of learning must be extended over a far wider range and in much greater detail than at present, and must reveal the similarities and differences between the many types and levels of modification of behavior before any attempt to formulate a general theory of learning will be of value.

REFERENCES

1. Beritoff, J. Ueber die individuell erworbene Tätigkeit den Nervensystem bei Tauben. *Pflüg. Arch. f. d. ges. Physiol.*, 1926, **213**, 370-406.
2. Bird, C. The effect of maturation upon the pecking instinct of chicks. *Ped. Sem.*, 1926, **33**, 212-234.
3. Bouman, L., & Grünbaum, A. A. Experimentelle-psychologische Untersuchungen zur Aphasie und Paraphasie. *Zsch. f. d. ges. Neur. u. Psychiat.*, 1925, **96**, 481-538.
4. Carmichael, L. A further study of the development of behavior in vertebrates experimentally removed from the influence of external stimulation. *Psychol. Rev.*, 1927, **34**, 34-47.
5. Cason, H. Criticisms of the laws of exercise and effect. *Psychol. Rev.*, 1924, **31**, 397-417.
6. ————. The physical basis of the conditioned response. *Amer. J. Psychol.*, 1925, **36**, 371-393.
7. Child, C. M. The origin and development of the nervous system. Chicago, 1921.
8. Detwiler, S. R. Experiments on the reversal of the spinal cord in Amblystoma embryos at the level of the anterior limb. *J. Exper. Zool.*, 1923, **38**, 293-321.
9. ————. Coordinated movements in supernumerary transplanted limbs. *J. Comp. Neur.*, 1925, **38**, 461-490.
10. Ebbecke, U. Die kortikalen Erregungen. Leipzig, 1919.
11. Franz, S. I. Variation in the distribution of the motor centers. *Psychol. Monog.*, 1915, **19**, 80-162.
12. Gelb, A., & Goldstein, K. Psychologische Analysen hirnpathologischer Fälle. Leipzig, 1920.

13. Goldstein, K. Das Wesen der amnetischen Aphasie. *Schweiz. Arch. f. Neur. u. Psychiat.*, 1924, **15**, 163-175.

14. Head, H. Aphasia and kindred disorders of speech. New York, 1926.

15. Herrick, C. J. Brains of rats and men. Chicago, 1926.

16. Holmes, G., & Lister, W. T. Disturbances of vision from cerebral lesions with special reference to the macula. *Brain*, 1916, **39**, 34-73.

17. Jennings, H. S. Modifiability in behavior. I. Behavior of sea anemones. *J. Exper. Zool.*, 1905, **2**, 447-472.

18. ————. Behavior of the starfish *Asterias forreri* de Loriol. *Univ. Calif. Publ. Zool.*, 1907, **4**, 53-185.

19. Johnson, H. M. A simpler principle of explanation of imaginative and ideational behavior and of learning. *J. Comp. Psychol.*, 1927, **7**, 187-235.

20. Kappers, C. U. A. Further contributions on neurobiotaxis. IX. An attempt to compare the phenomena of neurobiotaxis with other phenomena of taxis and tropism. *J. Comp. Neur.*, 1917, **27**, 261-298.

21. Klüver, H. An experimental study of the eidetic type. *Genet. Psychol. Monog.*, 1926, **1**, 71-230.

22. Koffka, K. The growth of the mind. New York, 1925.

23. Lapicque, L. Le poids du cerveau et l'intelligence. *J. de psychol.*, 1922, **19**, 5-23.

24. Lashley, K. S. Brain mechanisms and intelligence. Chicago, 1929.

25. Lashley, K. S., & Ball, J. Spinal conduction and kinaesthetic sensitivity in maze habit. *J. Comp. Psychol.*, 1929, **9**, 70-106.

26. Loeb, J. Comparative physiology of the brain and comparative psychology. New York, 1902.

27. Matthaei, R. Von den Theorien über eine allgemein-physiologische Grundläge des Gedächtnisses. *Zsch. f. allg. Physiol.*, 1921, **19**, Sammelref., 1-46.

28. Monakow, C. Z. Die Lokalization im Grosshirn. Wiesbaden, 1914.

29. Pavlov, I. P. Conditioned reflexes. Oxford, 1927.

30. Pechstein, L. A. Whole vs. part methods in motor learning. *Psychol. Monog.*, 1917, **33**, No. 99. Pp. 80.

31. Peterson, J. Completeness of response as an explanation principle in learning. *Psychol. Rev.*, 1916, **23**, 153-162.

32. Piéron, H. Thought and the brain. London, 1926.

33. Poppelreuter, W. Ueber den Versuch einer Revision der psychophysiologischen Lehre von der elementaren Assoziation und Reproduction. *Monatssch. f. Psychol. u. Psychiat. u. Neur.*, 1915, **37**, 278-323.

34. ————. Die psychischen Schädigungen durch Kopfschuss. Leipzig, 1917.

35. Stone, C. P. The congenital sexual behavior of the young male albino rat. *J. Comp. Psychol.*, 1922, **2**, 95-153.

36. Thorndike, E. L. Educational psychology. New York, 1913.

37. Thorsen, A. M. The relation of tongue movements to internal speech. *J. Exper. Psychol.*, 1925, **8**, 1-32.

38. Travis, L. E., & Hunter, T. A. The relation between 'intelligence' and reflex conduction rate. *J. Exper. Psychol.*, 1928, **11**, 342-354.

39. Turner, C. H. Behavior of the common roach (*Periplaneta orientalis* L.) on an open maze. *Biol. Bull.*, 1913, **25**, 348-365.

40. Watson, J. B. Behavior: an introduction to comparative psychology. New York, 1914.

41. Weiss, P. The relation between central and peripheral coordination. *J. Comp. Neur.*, 1926, **40**, 241-252.

42. Yerkes, R. M. The intelligence of earthworms. *J. Animal Behav.*, 1912, **2**, 332-352.

CHAPTER 15

LEARNING: II. EXPERIMENTAL STUDIES OF LEARNING

Walter S. Hunter
Clark University

The General Character of Studies of the Learning Process

Learning is one of the most characteristic aspects of human and infrahuman behavior. It is difficult in certain limiting conditions to distinguish between those changes in behavior which constitute genuine cases of learning and other changes in behavior which may be due to modifications of receptor function or to actual changes in the effectors, such as increased muscular strength. In general, however, we may say that learning is taking place wherever behavior shows a *progressive change,* or *trend,* with a repetition of the same stimulating situation and where the change cannot be accounted for on the basis of fatigue or of receptor and effector changes. Those trends in behavior which thus earn the title of learning are presumably due to changes in the nervous system and are possible because of the plasticity and retentivity of this system.

Studies of learning cover both the fact of acquisition and that of retention. Acquisition is itself cumulative and is to be understood only upon the assumption of the retention of the effects of previous stimulus-response activities. The evidences for retention and the methods of testing it, I have elsewhere (35, pp. 276-279) described as follows: "When we say that the co-ordination [of a stimulus-response activity] is retained, we mean one or all of three very concrete things: (1) The response can be aroused, or reinstated, by a presentation of the stimulus. Thus a subject who has learned to run a maze or to repeat the French equivalents of a list of English words may still be able to respond correctly to the respective stimuli after an interval of one week during which there has been no practice. (2) Although the subject may no longer be able to make the response to the stimulus as he has been trained to do, he may still be able to designate which stimuli he has been trained with and which stimuli he has not been trained with. In other words, he behaves differently to the maze or to the words on which he has been trained than he does to mazes or words on which he has not been trained. (3) Although the subject may no longer be able to make the response which he has been trained to make to the stimulus, and although he may react no differently to this stimulus, e.g., the list of English words, than to other comparable stimuli, e.g., other lists of English words, nevertheless his organism may still retain traces of of the results of previous training. We can establish this fact by having the subject relearn the original stimulus-response co-ordination. If we compare the amount of time required for the original learning with the time required for the relearning, we shall find, if there is any retention, that the relearning requires less time than did the original learning.

"The three conditions of behavior which constitute the evidence for the retention of stimulus-response co-ordinations also suggest the methods by which retention can be experimentally studied. The first method is that of recall or *reinstatement*. This method is used in all studies which involve recitation [the arousal of a response in the absence of its usual stimuli] or other form of recall of the specific co-ordination under investigation. A specific form of this method is that of *paired associates,* devised by Calkins [11] and perfected by Müller and Pilzecker [60]. In this form of the method the material to be used as stimuli is presented in pairs (of words, pictures, etc.). The subject is instructed to speak the words, *e.g.,* and the entire series of paired stimuli are presented serially a specified number of times. Then the first members of the pairs are presented to the subject in a changed order. He is now instructed to recall, *i.e.,* to reinstate, the response formerly aroused by the second member of the pair. Retention is measured by the length of time that it takes him to reinstate the required response and by the proportion of right and wrong responses. Thus the series *a-b, c-d, e-f, g-h* is presented to the subject one pair at a time, each pair for one second. After the fifth reading, let us say, the subject is then shown *c, g, a* and *e,* and instructed to reinstate the responses (*d, h, b, f*) which had formerly been given to the absent stimulus of the pair.

"The second method of studying retention is usually called the method of recognition. This method requires the subject to discriminate between old and new stimuli, *i.e.,* between stimuli with which the subject has been trained and stimuli with which he has not been trained. We shall therefore call this second method the *age of stimulus method.* A typical procedure in the use of this method is as follows: (1) Colors, grays, odors, tones, pictures, advertisements, words, whatever it be, the stimuli for the experiment are chosen. (2) This selected material is now presented to the subject either serially or in pairs. During the presentation the subject is instructed to respond with the verbal response which he has learned to call the name of the stimulus, or he is permitted to make no overt response. (3) The stimuli may be presented any number of times up to the point where the subject can reinstate all of the responses which he has learned for the series. And (4) at the end of any presentation the stimuli may be presented to the subject in the same arrangement in which they were originally or in a different one, but in either case new stimuli will be mixed in with the old. The subject is now instructed to make one type of response to the old stimuli and another to the new. This discriminative test need not be given immediately after training. It may be deferred for any length of time chosen by the experimenter.

"The third method of studying retention is *the saving method* of Ebbinghaus [21]. In this method we record the number of trials, the amount of time, and the number of errors in the learning process for any given stimulus-response co-ordination. Later, after the lapse of a certain length of time, the subject again learns the co-ordination up to the same level of skill required in the original learning. The difference between the records for the original learning and the same measurements for the relearning is an indication of the amount of retention. If a subject requires as much time for the second learning as for the original learning, there is no evidence that he has retained any of the initial practice effects."

The vast range of experimentation and discussion in the field of the learning process precludes the possibility of our covering the entire field in the present chapter. Learning is intimately involved in the development of all behavior patterns. This is clearly indicated in the field of reflexes by Pavlov's (65) work on the conditioned reflexes and in the field of instinct by the work of Shepard and Breed (85), Moseley (59), Bird (3), and Carmichael (12) on the possible rôle of maturation in the development of behavior. We shall discuss one aspect of the maturation problem when we consider the relation of age to learning capacity; but the aspect raised by the studies just mentioned will not concern us because the maturation of the specific stimulus-response mechanisms involved in experiments upon learning has apparently been completed before the subject reaches the experiment. Learning is also intimately related to the problem of work, *i.e.*, to the problem of the exercise of forms of response, and to that of thinking behavior where the acquisition and use of symbolic processes is of characteristic importance. The limitations of space will prevent our developing these topics. Furthermore, we can only mention the importance of a study of the learning process for the adequate understanding of those topics which the psychologist has labeled attention and testimony. Thus studies on the span of attention, the immediate memory span, and testimony are all alike in that they require the subject to reinstate certain learned responses after a single presentation of the stimulus. They differ only in the temporal duration of the stimulus. If the stimulus is presented for 1/50 of a second, the experiment is classified as one on attention; whereas with longer exposure times, the behavior is classified as testimony or memory.

Since the problem of acquisition and retention concerns stimulus-response co-ordinations in an individual subject, the complete experimental record of the problem involves a study of the effect upon acquisition and retention of: (1) variations in the stimulating conditions, such as temporal duration, frequency, whole vs. partial presentation, and amount; and (2) variations in the condition of the subject, such as fatigue, age, sex, cultural status, ability, incentive, neural condition, and character, amount, and recency of past training. In the present chapter we shall discuss certain of the more important problems here suggested, stressing constantly the analysis of experimental findings. The references at the close of the chapter will afford the reader not only an opportunity to examine the original experimental material referred to in this chapter but to extend his reading to other experiments which it was not possible to analyze. Let us turn first, then, to a study of the curve of acquisition.

The Learning Curve

The learning curve is a graphic representation of the change in behavior which is brought about by successive periods of practice all of which are represented on the curve. In contrast to this is the sample performance curve illustrated in Figure 3. In this graph improvement is not the result of the practice represented by the scores on the curve but is a result of the practice which occurs outside of the experimental period. No particular difficulty arises in the construction of either type of curve for an individual subject. In such a case values along the abscissa represent successive practice (or test) periods equal in magnitude one to the other, and the values along the ordinate represent some quantitative aspect of behavior such as the number of errors, the number of correct responses, or the time consumed per unit on the abscissa. Curves for individual subjects usually fluctuate considerably from practice period to practice period, not because learning is an intrinsically fluctuating process but because the experimenter does not sufficiently control the experimental conditions under which the learning takes place. It is furthermore to be noted that any particular learning curve can represent but one aspect of the trend in behavior and that it may not faithfully and accurately represent even this aspect. Thus progress may be rapid in the error aspect of the behavior and relatively slow in the time aspect; and when progress is slow, as shown in a plateau, the lack of apparent improvement may only be due to the crudeness of the measures employed.

We have spoken so far only of learning curves for individual subjects. The construction of curves which shall represent the learning process in groups of subjects is not so simple. The learning process extends from the beginning of practice until some arbitrarily established degree of skill is attained, e.g., three perfect trials in succession in the maze. Subjects will differ in the number of trials required for mastery of the problem. If, therefore, all subjects are given the same number of trials, let us say 30, and if the curve is constructed by averaging the records of all subjects for each successive trial, the resulting graph will represent how the average performance changes from the first to the 30th trial. (See Figure 13.) It will not represent the progress made in learning the problem set, for some of the subjects will have mastered the problem in much less than 30 trials, or practice periods, and hence will be greatly overtrained. The conventional method of constructing a curve which will portray the learning process in a group of subjects is illustrated below in Table 1. Vincent (108, pp. 16-17) has suggested the best method of constructing a curve which throughout its extent will be representative of the whole group of subjects and which will portray the change of behavior from the beginning of practice until the problem is mastered.

TABLE 1

ERROR RECORDS FOR FIVE RATS IN LEARNING A SIMPLE MAZE

The table shows two conventional methods of obtaining averages

	Rat 1	2	3	4	5		
Trial			Errors			Av. error	Av. error
1	40	20	30	25	10	25	25
2	25	30	24	10	12	20.2	20.2
3	8	20	18	7	6	11.8	11.8
4	2	15	10	4	3	6.8	6.8
5	1	10	5	1	5	4.4	4.4
6	0	3	0	1	2	1.2	2.
7	0	2	3	0	3	1.6	3.6
8	0	1	0	1	2	.8	1.3
9	0	1	0	0	1	.4	1.
10	0	0	0	0	1	.2	1.
11	0	0	0	0	2	.4	2.
12	0	1	0	0	1	.4	1.
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0

The construction of a Vincent learning curve can be illustrated from the data of Table 1, which contains the error records for five rats in a simple maze. If three perfect records in succession be taken as the criterion of mastery, learning was completed as follows: rat No. 1, 5 trials; rat No. 2, 12 trials; rat No. 3, 7 trials; rat No. 4, 8 trials; and rat No. 5, 12 trials. If we average the material by trials as is done in either of the last two columns of table, we either include some records beyond the completion of learning or we omit some of the subjects. The Vincent curve, which is not open to these criticisms, is constructed by dividing the learning periods into an equal number of parts on the basis of the number of trials required for mastery. These parts are then averaged as a basis for the graph. Suppose we divide the learning of each subject into ten equal parts. Rat No. 1, who learned in 5 trials, learned in ten parts each of which is one-half a trial. The ten parts of No. 2's learning are each 1.2 trials long. When the number of errors in each tenth are calculated and averaged for the group. the values of Table 2 result.[1]

An excellent illustration of the practical value of the Vincent type curve is to be found in the studies by Kjerstad (43), Robinson and Darrow (81), and Robinson and Heron (82), where the curves of learning for materials of different lengths are compared. Figure 1 from Robinson and Darrow plots, in the conventional fashion, the average amount learned against the learning time. It would appear from these curves that learning progresses more rapidly in the initial stages as the length of the material to be learned decreases. Such a conclusion, however, would be unjustified because the learning process is more condensed in the shorter

---

[1] Kjerstad (43, p. 26) describes a graphic method of arriving at the values for the Vincent curve.

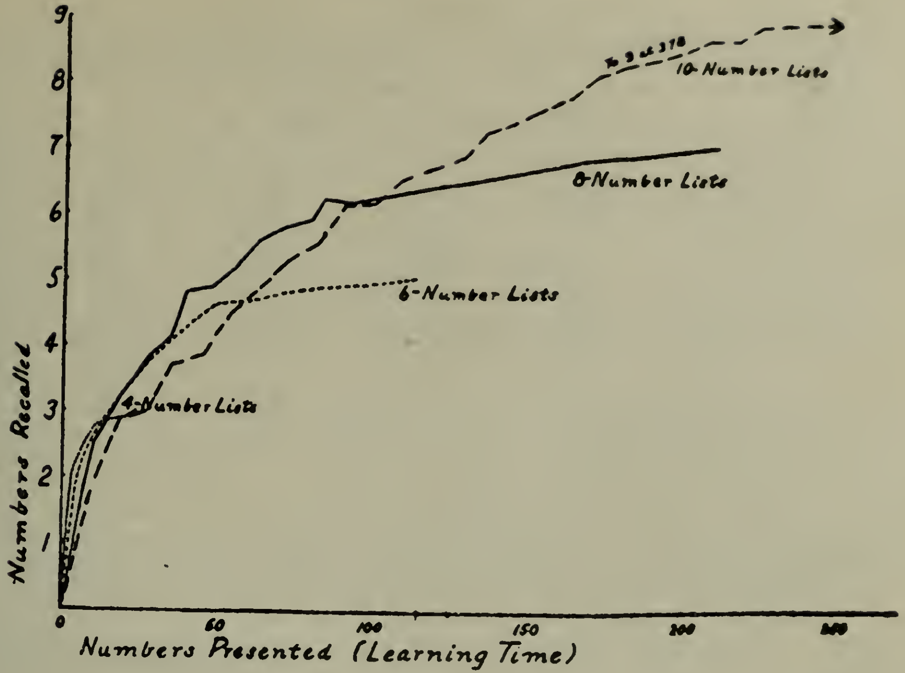


FIGURE 1

Curves for Different Amounts of Material, Showing the Relation between Amount Recalled and Learning Time

(From Robinson and Darrow, *Amer. J. Psychol.*, 1924, **35**.)

curves than in the longer ones. In order to compare progress in the various problems, curves must be constructed which will be of equal length. When the results are averaged by the Vincent method, the authors secure the data of Table 3, which shows that the progress of learning is essentially the same for all lengths of material, the shortest list of numbers and of syllables offering the only discrepancy. Robinson and Darrow thus confirm the prior findings of Kjerstad that the form of the learning curve is independent of the amount of material learned.

Learning curves have one of two general forms depending upon the values that are plotted on the ordinate. If correct responses are recorded on the ordinate, the resulting curve will be some variant of the type shown in Figures 1 and 3 because the number of correct responses increases with the increase in practice. When, however, time or error records are plotted on the ordinate, the curve is of the general character of Figure 7. Aside from this ascending or descending characteristic of the curve of learning, attention should be directed to changes in the form of the curve which result from changing rates of the acquisition or elimination of responses. We shall reserve for following sections the discussion of long periods of little or no change in the curve, *the plateaus,* and the discussion of sudden improvements in the subject's performance which have been accredited to *insight.* At present we shall confine ourselves to comments upon those changes in the curve which occur at the beginning and end of learning.

TABLE 2

RECORDS FOR A VINCENT CURVE BASED ON THE DATA OF TABLE 1

Tenths	Rat 1	2	3	4	5	Av.
	Errors in successive tenths					
1	20	26	21	20	12.4	19.8
2	20	32	18.6	11	12	18.7
3	12.5	21	16.2	6.8	5.4	12.3
4	12.5	14	12.6	5	5.2	9.8
5	4.	5	8.6	3.2	3.	4.7
6	4.	2.2	9.	.8	3.4	3.8
7	1.	1.2	3.5	.8	2.	1.3
8	1.	.6	.5	.4	1.2	.7
9	.5	0.0	.9	.2	2.	.6
10	.5	1.	2.1	.8	1.4	1.1

TABLE 3

THE RELATIONSHIP BETWEEN AMOUNT OF MATERIAL TO BE LEARNED AND THE PROGRESS OF LEARNING AS REVEALED BY THE PERCENTAGES OF MATERIAL LEARNED IN SUCCESSIVE SIXTHS OF THE LEARNING PROCESS
(After Robinson and Darrow)

Per cent of total learning time	Per cent of materials mastered (three-place numbers) for lists of different lengths 4 nos.	6 nos.	8 nos.	10 nos.
16.7	19.4	31.8	29.3	30.4
33.3	36.1	44.5	45.9	45.3
50.	49.8	54.0	55.3	54.6
66.7	63.1	62.7	64.0	64.7
83.3	82.3	71.0	75.8	70.4
100.	100.	100.	100.	100.

	Per cent mastered for nonsense syllable lists of different lengths 6 syl.	9 syl.	12 syl.	15 syl.	18 syl.
16.7	23.3	33.9	34.6	30.9	33.1
33.3	45.	51.2	52.3	52.2	53.2
50.	60.5	62.	66.3	62.0	66.3
66.7	72.3	73.6	74.9	74.2	76.3
83.3	83.4	80.8	80.5	80.3	82.8
100.	100.	100.	100.	100.	100.

The initial period of learning as portrayed by time and error curves is usually one of rapid change as indicated in Figures 1 and 7. Such a change indicates either that the factors, e.g., emotional disturbances, making for high time and error scores quickly disappear as the subject becomes adjusted to the experimental situation or that certain easily established components of the final habit have been acquired. Figure 9 shows an error curve which remains on the same level for several trials and then drops. In the absence of an experimental analysis of the situation from which the curve is taken, we must content ourselves with the hypothesis that an adjustment must be made to certain factors which are difficult to master before learning can proceed. This difficulty of the initial aspects of the problem would be adequate to account for the curve. Curves drawn for correct responses may also show initial periods either of rapid change, Figure 1, or of slow change, Figure 2. Peterson's (70) study of ball-tossing gives us this latter curve. *Ra* represents the average number of successive catches per practice period. The curve is concave to the ordinate, as are also curves constructed by Swift (94) and Batson (2) for the same habit, thus indicating that the early adjustments necessary to score successive catches are acquired with difficulty. After the preliminary stages of learning, progress becomes more rapid until the final level of performance is reached.

If practice is continued long enough, all curves of learning reach a final level of attainment. Probably all curves reach this level through a period of negative acceleration. Peterson's curve shows a final level which is attained suddenly. This, however, is an artifact due to the limitation of each practice period to 200 catches and to the consequent fact that not more than 200 successive catches could be made per practice period. If the subject had been instructed to increase his number of successive catches as much as possible, the resulting curve would undoubtedly have leveled off gradually rather than sharply. The final negative acceleration of learn-

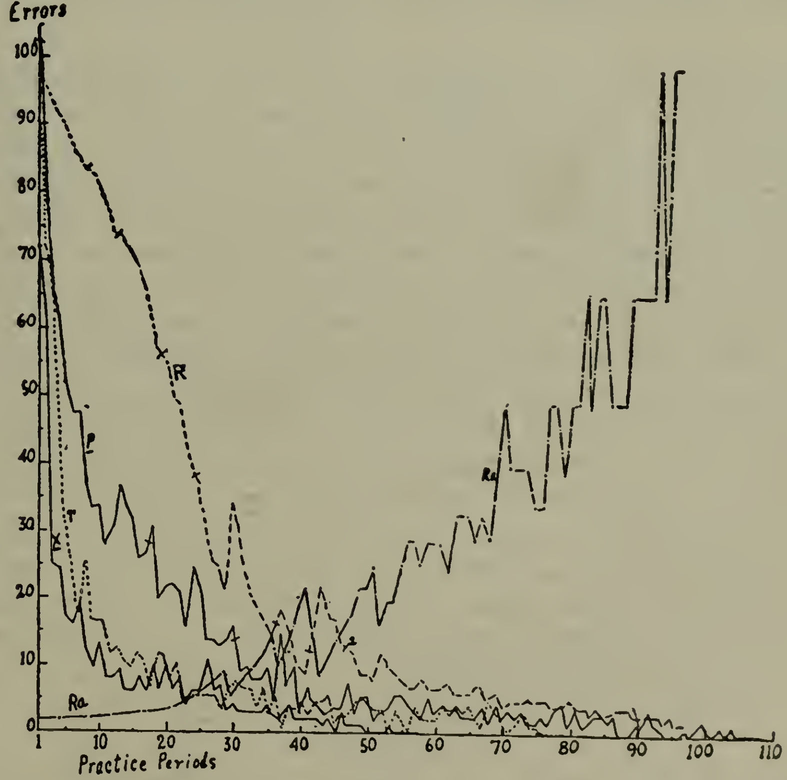


FIGURE 2

Curves for the Elimination of Errors and for Achievement (*Ra*) in Ball-Tossing

(From Peterson, *J. Exper. Psychol.*, 1917, 2.)

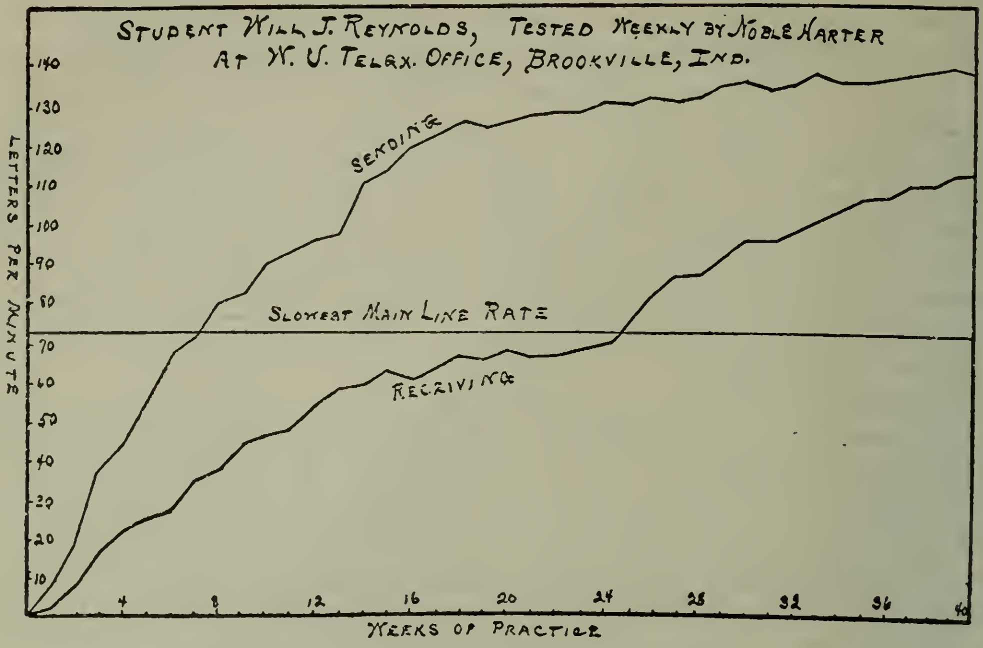


FIGURE 3
Sample Performance Curves for Sending and Receiving Telegraphic Messages
(From Bryan and Harter, *Psychol. Rev.*, 1899, **6**.)

ing curves may be due to the difficulty in mastering the final aspects of the habit either because the most difficult components tend to be learned last[2] or because the integration of the total series of responses is itself an accomplishment difficult to attain. Thus the individual stimulus-response components may be mastered, but their appearance in the proper sequence and in the minimum time may yet constitute a problem for mastery.

Whenever a problem is set for the subject in such a form that small differences in the accuracy of responses may be measured and whenever speed is a factor in the behavior, the physiological limit of the organism sets the limit beyond which the subject's performance cannot go. The physiological limit is rarely if ever reached. The subject's performance rather ends on a level determined by the incentive under which he works as well as by the difficulty of making additional improvement. Whether the activity engaged in be golf, typesetting, or addition, the subject tends to remain on a final level of attainment which is socially and economically acceptable. Even though practice is long continued, no improvement may result unless new incentives, such as higher wages or personal competition, are added to the working conditions.

PLATEAUS

A plateau in the learning curve is a period during which the subject makes little or no progress toward mastery of the problem. Bryan and

---

[2] On this point see Peterson, G. M. (69).

Harter (10) were the first to emphasize the problem presented by the occurrence of plateaus and to seek for their explanation. Figure 3 presents two curves, one for sending and the other for receiving telegraphic messages. These curves are sample performance curves, which indicate the degree of skill attained by the subject after the indicated weeks of practice. Once each week the subject was tested and the number of words (selected from standard sentences) sent or received during several two-minute intervals were averaged and translated into letters per minute to give the values entered on the ordinate. Figure 4, a receiving curve, was constructed in the same way for a subject who had studied telegraphy about six weeks before beginning the experiment. This curve shows the progress made in the ability to receive letters not making words, words not making sentences, and words which did make sentences.

Bryan and Harter explain the plateau of the receiving curve in terms of the acquisition of a hierarchy of verbal habits. "The curves of figure [4] show also, however, that for many months the chief gain is in the letter and word habits, that the rate of receiving sentences is, in this period, mainly determined by the rate of receiving letters and words, and that rapid gain in the higher language habits does not begin until letter and word habits are well fixed" (10, p. 356). The sending curve shows no plateau "because, as in the early part of the receiving curve, the various habits are acquired simultaneously."

Swift (94, 95, 96) in his studies of ball-tossing, typewriting, and learning Russian and Book (4) in his study of typewriting found plateaus of greater or lesser length. Swift believes that the plateaus are partly due to failure in the subject's effort and enthusiasm, partly to the experimenter's inability to measure the progress that is taking place, and partly to the necessity for the subject to perfect certain responses before he can proceed to a higher level of efficiency. One of Book's figures is reproduced in our Figure 5. The curves represent the number of strokes per minute on successive days of practice in copying material by the sight method of typewriting. Book partially agrees with Swift that the plateaus are due to a relaxation of effort on the part of the subject rather than to the influence of a hierarchy of habits, but he adds to the former factor that of excessive effort wrongly applied.

Batson (2), who has made the most extended study of plateaus since the work by Bryan and Harter, trained his subjects in tossing and catching two balls. In order to perform such a task accurately, it is necessary to throw the ball in the proper direction, to throw it with the proper force, and to time the opening and closing of the hand accurately enough to catch the ball. Batson devised tests in which the subject must react primarily to directional, distance (force), and time factors. The directional test involved rolling a steel ball along a smooth board in an attempt to strike a small wooden block. The force, or distance, test required the subject to roll a rubber ball so that it would stop as near to a given mark as possible. The timing test required the subject to release a small lead

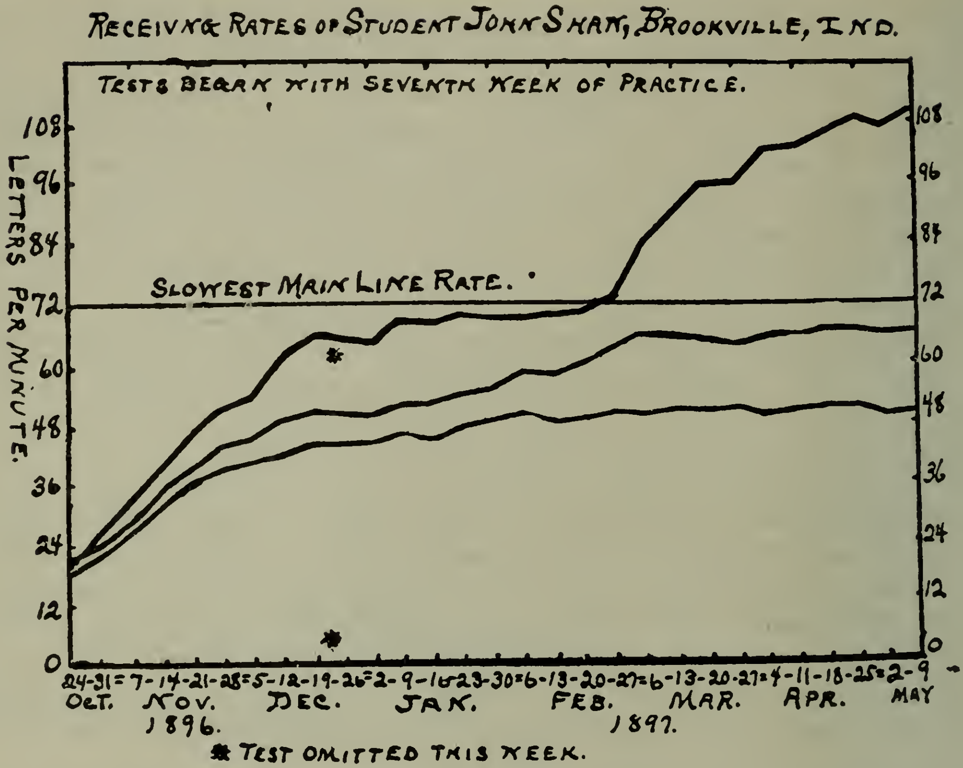


FIGURE 4

Sample Performance Curves for Receiving Telegraphic Messages

(From Bryan and Harter, *Psychol. Rev.*, 1899, 6.)

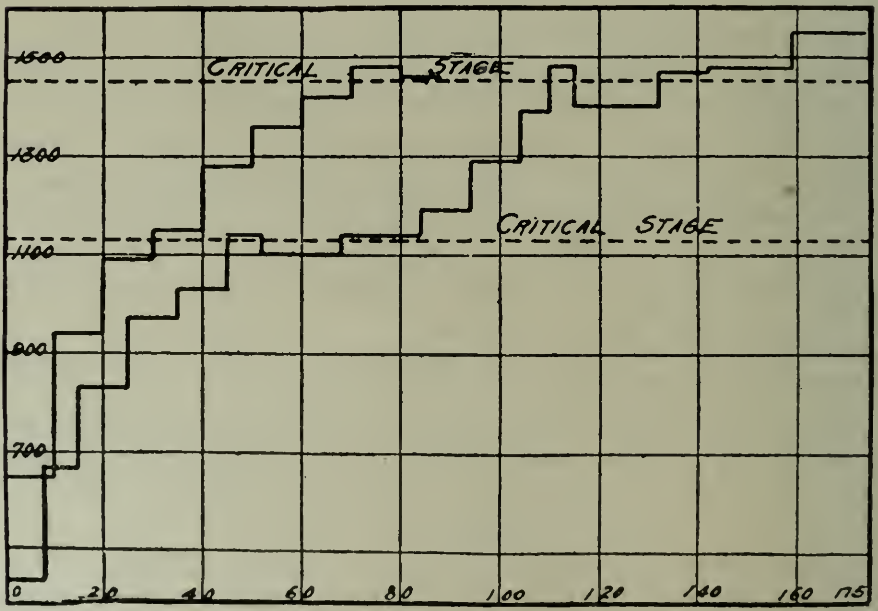


FIGURE 5

Curve for Learning Typewriting

(From Book, W. F. The psycholgoy of skill. New York: Gregg, 1925. Reproduced with the permission of the Gregg Publishing Co.)

shot so that it would roll down a groove and reach the end at the proper time to fall into the pocket of a disc which rotated at a constant speed past the end of the groove. Since these three analytical learning tests were not given simultaneously with the ball-tossing test which they were designed to analyze, it is impossible to correlate progress in ball-tossing with improvement in the reactions to direction, distance, and time as it is possible in Figure 4 to compare the progress in receiving letters and words with the progress in receiving connected discourse. In any case it is highly questionable whether time, distance, and direction are involved in Batson's analytical tests in the way that they are in ball-tossing. One could hardly look, in that case, to these tests to clarify the curve for ball-tossing. Batson concludes that plateaus do not occur in simple types of learning (and yet his Figure 16*a,* presenting the score in the direction test, shows a well-marked plateau), and that they may or may not occur in complex learning.

The whole problem of plateaus is greatly in need of experimental clarification. Since the plateau is a period of little or no progress in learning, any factor which retards learning may theoretically produce plateaus. Lack of effort, misdirected effort, lack of incentive, emotional upset, fatigue, habit interference, and the presence of unusually difficult components of the problem to be learned are some of these factors. Trow and Sears (106) have recently presented evidence indicating the possible rôle of habit interference in the production of plateaus; and Hunter, in some work as yet unpublished, has secured maze curves for rats which show marked plateaus due to the great difficulty in eliminating certain errors. Before we can be certain that hierarchies of habits are concerned in the production of plateaus, we must have additional experiments like those of Bryan and Harter where the external and internal conditions of work are carefully controlled. Plateaus may be a genuine and necessary aspect of certain processes of learning and in other instances they may merely indicate the experimenter's failure to keep working conditions constant for his subject.

The Nature of Insight

In the topic of insight we not only extend our consideration of the learning curve to include the problem of a sudden improvement in efficiency, but we also face the theory that there is present in some learning a factor, designated as insight, which is not of the same order as the factors involved in the conditioning of reflexes or in so-called trial-and-error learning. Our general conclusion will be that there is no adequate evidence for the assumption of a peculiar insight factor or for the assumption that there is more than one kind of learning. Let us begin with a consideration of Köhler's work on apes.

Köhler, (46) in his extension of Hobhouse's (30) type of animal experimentation, defends the thesis that insight plays a rôle in the learning process. Although Köhler gives no analysis of the nature of insight, he contrasts his own experiments with those of Thorndike (97) and hence may be said to treat insight as a form of intelligence. Intelligent solutions of problems, from Köhler's point of view, are to be contrasted with accidental solutions as follows: "The genuine achievement takes place as a single continuous occurrence, a unity, as it were, in space as well as in time; in our example, as one continuous run without a second's stop, right up to the objective. A successful chance solution consists of an agglomeration of separate movements, which start, finish, start again, remain independent of one another in direction and speed, and only in a geometrical summation start at the starting point, and finish at the objective" (46, pp. 16-7). Successful chance solutions would be illustrated by the behavior of a subject in his first trial in a maze where almost every possible error is made and where the exit is merely blundered upon. Another illustration of the same phenomenon is found in the behavior of Thorndike's cats (97), who were confined in a box from which they could escape only after having turned a latch. At first the cats attacked every portion of the cage and effected their escape only after having turned the latch in the process of this random scrambling. In contrast with these chance solutions, Köhler emphasizes that genuine achievement, i.e., intelligent behavior (or behavior with insight), involves but little if any useless activity, that it occurs as a single continuous and integrated response. In order that such behavior may be manifested Köhler says that all aspects which are essential to the solution of the problem must be simultaneously present to the subject so that he can obtain a view of the entire situation in all of its relevant aspects. This is equivalent to saying that all of the stimuli whose action is necessary for the correct response must be simultaneously present. Such a condition is fulfilled only with visual stimuli, and the result is that insight would thus be limited to visually determined behavior. This in itself leads one to question the hypothesis, for it is highly doubtful that so important a factor as insight is alleged to be would be limited in this manner. However, the question is one of fact to be determined only by experiment.

In order that the reader may have a concrete example from Köhler of intelligent, or insight, behavior, I quote the following: "A species of reversal of the 'tool-using' experiments consists in placing a movable object across the path of the objective so that the problem can be solved only by its *removal*: . . .The box [which was the obstacle] was placed in the barred room in immediate contact with the bars and standing on the smaller end so that it could easily be knocked over. Outside the bars, and immediately opposite the center of the box, the bananas lay on the ground; they could be reached at once with a stick, if the box were pushed aside or even knocked over . . . She [Tschego] crouched beside the box, facing the bars. For some time nothing happened. Then, however, some of the smaller apes approached from outside the cage . . . Each time, though, Tschego repulsed them with threatening gestures . . . The youngsters finally gathered closely around the fruit, but the danger inspired Tschego; she gripped the box, which was like a toy in her arms, jerked it backwards, stepped up to the bars, and took the fruit . . . The 'obstacle' test was *not* solved . . . by a series of imperceptible pushes involuntarily given to the [box] in the act of stretching towards the prize. Quite the contrary: *during the lapse of two hours, Tschego did not move the* [*box*] *one millimeter from its original position,* and when the solution arrived, the [box] was not *shouldered* to one side, but *suddenly gripped* with both hands, and thrust back. It was a *genuine* solution" (46, pp. 61-5).

With these illustrations before us, we may proceed with our analysis. There are two ways of defining insight: one method is to make insight a purely descriptive term referring to a certain type of behavior; the other is to make it an explanatory term referring to something (usually psychic) which influences and controls behavior. Köhler has usually adopted the first course although much of his discussion involves the second. Other investigators, notably Koffka (45) and Yerkes (122, 123), who discuss insight also use first one and then the other meaning of the term. Thus Koffka writes: "It has been shown that improvement in efficiency goes hand in hand with an increased insight into the nature of the task. We use this word, insight, without theoretical presuppositions, in the common sense in which everyone takes it. If one knows that he is to remove a ring in a certain puzzle, and that in order to do so he must first move this piece and then that, turn the puzzle over and do something else, his procedure will be said to possess a greater degree of insight than the procedure of another person who simply goes ahead without any plan at all" (45, p. 179). And again, "The sudden grasping of the solution which results is a process that runs its course in accordance with the nature of the situation, so that the complete solution of the problem takes place with reference to the configuration of the field of perception; and this is what Köhler maintains to be the criterion of insight" (45, p. 214). Yerkes' use of the term in an explanatory way is indicated by the following quotation: "Insight is used throughout this report to designate varieties of experience which in us are accompaniments of sudden, effective, individually wrought adaptations to more or less distinctly new and problematic situations" (123, p. 155). It is impossible to do more than philosophize when dealing with such definitions of insight. The only practical and scientific problem is this: Are there types of behavior or characteristics of behavior so different from those present in the formation and use of habit in general that we are justified in speaking of insightful behavior?

Insight is described by Köhler as a form of response which appears suddenly and runs its course smoothly from the beginning until the attainment of the goal, under conditions where all of the essential aspects of the problem can be viewed at one time by the subject. We have already rejected this last characteristic because it seeks to confine the alleged phenomenon to visually controlled behavior. The "smooth, continuous" aspect of behavior is characteristic of all well-established habits, including the maze habit, and so cannot serve to mark off one habit from another. (Köhler did not know the past history of his apes, and so could not know to what extent the "solutions" were merely reinstated habits.) This leaves us with the possibility that behavior which appears suddenly is insight. We shall discuss this after noting the characteristics proposed by Yerkes. Yerkes writes as follows: "In acts which by us are performed with insight or understanding of relations of means to ends, we are familiar with certain characteristics which are important if not also differential. The following is a partial list of features of such behavior. . . .(1) Survey, inspection, or persistent examination of problematic situation. (2) Hesitation, pause, attitude of concentrated attention. (3) Trial of more or less adequate mode of response. (4) In case initial mode of response proves inadequate, trial of some other mode of response, the transition from one method to the other being sharp and often sudden. (5) Persistent or frequently recurrent attention to the objective or goal and motivation thereby. (6) Appearance of critical point at which the organism suddenly, directly, and definitely performs required adaptive act. (7) Ready repetition of adaptive response after once performed. (8) Notable ability to discover and attend to the essential aspect or relation in the problematic situation and to neglect, relatively, variations in non-essentials" (123, p. 156).

Yerkes' list of characteristics is proposed as a basis for the identification of a specific form of behavior called insight. And yet nos. 1, 2, 3, 4, 5, and possibly 8 are features of all learning processes whether they involve rats or men and whether the material to be learned is a maze, a brightness discrimination, or a list of nonsense syllables. Furthermore, all of the eight features occur where learning is completed in one trial. Indeed the eight features are all revealed in the behavior of the unicellular organism *Stentor* as described by Jennings (38, chap. 10). A child who refuses bitter medicine after one trial, or a chick who after one trial refuses to peck at certain food would qualify for insight behavior, and yet the well-established principle of conditioned reflexes makes the assumption of insight unnecessary for the explanation of the behavior. As in Köhler's statement so here the one feature that calls for extended discussion is that of the sudden drop in the learning curve.

Figures 6-9 present certain learning curves relevant to the problem of insight. Figure 6, presented by Thorndike (97, p. 41), represents the change in time during 13 trials in which a cat was trained to escape from a latch box. Figure 7 is a curve from Watson (110) showing the time records of 19 rats in a maze. Figure 8 is a curve representing the work of an ape in a multiple-choice problem studied by Yerkes (122, p. 27). Figure 9 reproduces curves from Clements' (16) study of the effect of different lengths of delay upon the learning by rats of the shorter of two paths each leading to a food box. All of these curves show sudden drops. Are we, therefore, to conclude that insight was present in all instances? How sudden must a drop be before insight behavior is involved? Or how much of a drop must be present? If a complete elimination of error within one or two trials shows insight, does a partial elimination show partial insight? There seems to be no proper objection to the affirmative answer to this last question, and yet

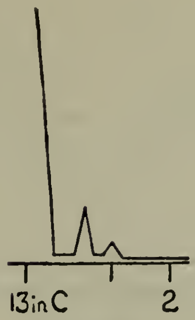


FIGURE 6

A Curve Showing the Decrease in Time Required by a Cat to Escape from a Problem Box
(From Thorndike, *Psychol. Monog.*, 1898, **2**.)

such an answer makes any inference from drops in the learning curve to the presence of insight seem futile, for all learning curves are composed of a series of drops. The most pertinent and practical question to raise concerning such changes in the curve of learning is not "Does such a curve reveal the presence of insight?" but "What factors cause such drops in the learning curve? This latter question is quite within the range of experimental solution. For example, the sudden drop in the maze curve, Figure 7, is largely due to the adaptation of the subject to the experimental situation and to the establishment of behavior oriented toward the exit of the maze. Undoubtedly other factors exist which are equally susceptible to a concrete determination. No experiments have been conducted to determine why the sudden changes occur

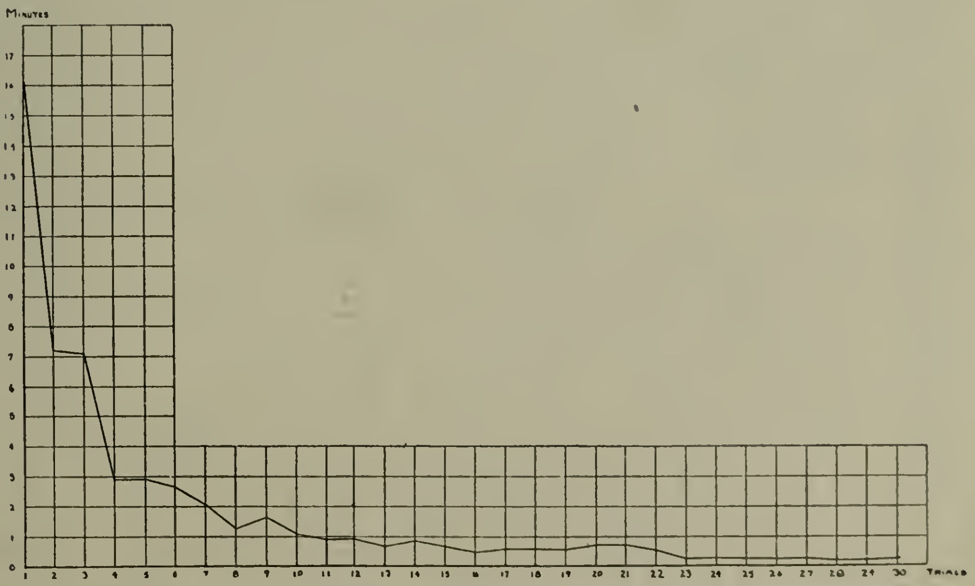


FIGURE 7

A Curve Showing the Average Time Required by 19 Rats in Learning a Maze
(From Watson, *Psychol. Monog.*, 1907, **8**.)

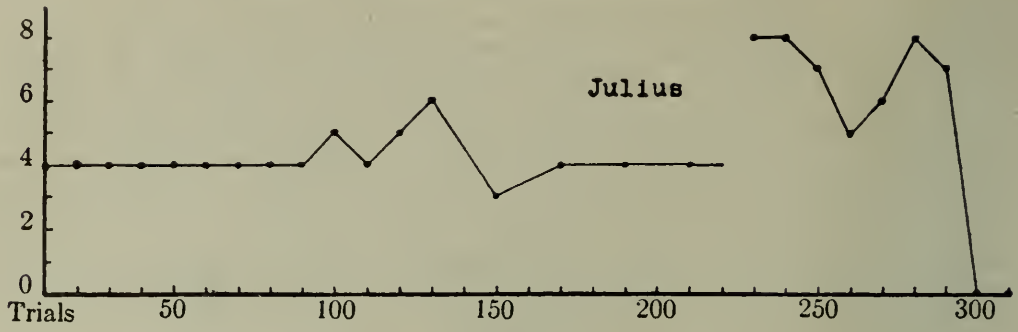


FIGURE 8

A Curve Showing the Progress Made by an Ape Toward the Mastery of a Multiple-Choice Problem

(From Yerkes, *Behav. Monog.*, 1916, **3**.)

in the curves of Figure 6, 8, and 9. However, Figure 6 indicates learning in one trial, and the experimental results on conditioned reflexes indicate that the phenomenon may well be due to the conditioning of the non-effective stimulus by some other stimulus which has a pronounced "right of way." Figure 8 suggests very strongly either (1) that certain factors blocked the appearance of learning so that our crude methods did not reveal what was taking place until suddenly the correct response appeared, or (2) that, as probably in Köhler's work, the sudden appearance of the correct response was essentially a reinstatement of previously learned behavior. The two curves in the middle of Figure 9 were clearly derived from responses which took place under conditions which

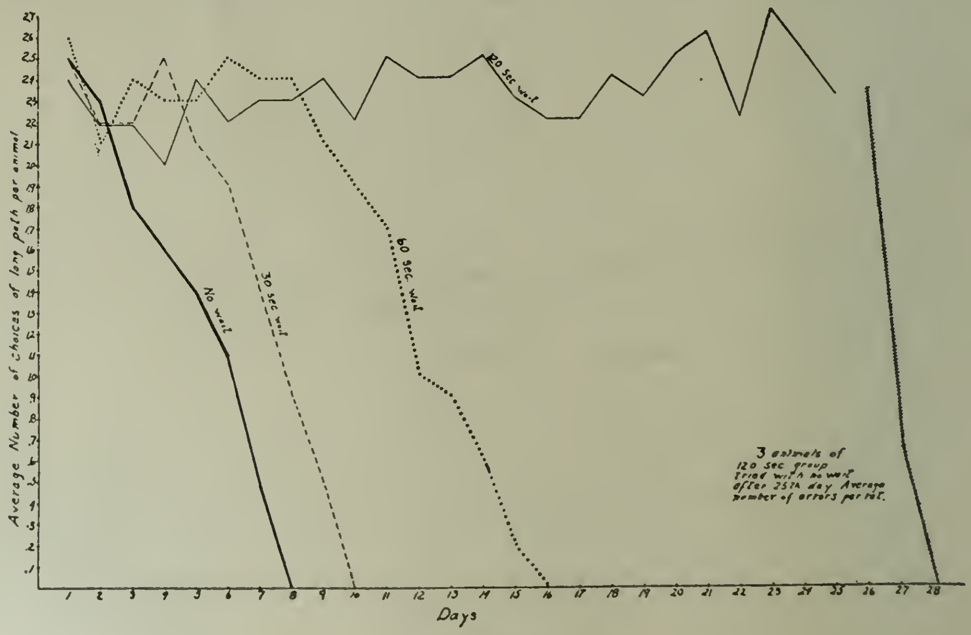


FIGURE 9

Learning Curves for Rats

The figure is described in the text. (From Clements, *J. Comp. Psychol.*, 1928, **8**.)

for a time failed to produce a measured degree of learning, but just what these detrimental conditions were has not been determined. (It should be noted that curves 1 and 4 of this figure were not secured under conditions comparable with those which produced curves 2 and 3.) We are thus forced to the conclusion that no causal factor of insight has been demonstrated and that to resort to such a hypothesis is to desert the straight-forward experimental attack upon the determining conditions of behavior from which alone an adequate solution can be derived. All learning is essentially of a kind, the modification of behavior as a result of repeated stimulation under specified conditions. Some learning periods may be shorter than others, or may result in a higher degree of efficiency than others, or may involve verbal guidance, but there is no adequate reason for believing in two kinds of learning.

We turn now to a consideration of two important ways in which the stimuli involved in learning may be presented to the subject: one involves the whole vs. the part method of presenting the stimuli which are to be responded to, and the other concerns the frequency with which the stimuli are given.

Whole versus Part Methods of Learning

If a subject is to learn a sequence of responses such as reciting a poem or running a maze, will it be more economical to repeat the entire sequence of responses at each trial or should the problem be broken up into parts and the parts learned separately? What will be the relative merits of the two methods when they are evaluated in terms of retention? These problems have been extensively investigated and discussed since Steffens' work in 1900 (91). In general, the results have indicated that the whole method is more economical than the part method both in learning and in retention. Exceptions, however, have been found by those who favor this conclusion, and Reed (78), Winch (118), and Pechstein (67) have presented definite experimental evidence in favor of the part method. Let us, therefore, examine briefly some of the experiments bearing upon the problem.

Steffens' work was carried out with poetry and nonsense syllables. The subjects, when allowed to study the material without definite instructions, followed a part method in which they studied a few lines, then added a few more lines, and repeated the first ones. Difficult passages were frequently and slowly read. When this normal method was compared with the whole method, the former required an average of 9 per cent more time for learning than the latter. With nonsense syllables, the normal method was superior for the first 10 days' work; but during the last 24 days this was reversed. An experiment was conducted where the subject learned three eight-line stanzas daily. One half of a stanza was mastered, then the second half, and finally the whole was read and mastered. On the average this method was inferior to the whole method by 63.6 secs. and by 1.6 trials. The work continued for 16 days. On only one of these days was the part method superior. When the stimulus material was presented on a rotating drum so that the exposure time during learning was the same for both methods, the whole method had an advantage of three trials.

Larguier des Bancels (51) confirmed Steffens' work and added the fact that the whole method of learning is more favorable for the retention of poetry than is the part method. This finding for retention was confirmed by Pentschew (68) who found, however, that with children the part method was slightly superior both in learning and relearning when nonsense syllables were used. Pentschew also found that where the whole method was superior in number of repetitions required the part method often had the advantage from the standpoint of time consumed. Pyle and Snyder (76) extended the problem to longer selections of poetry and report as follows: ". . . .without any exception, the method of the whole is more economical than is the part-method, and . . . .the saving is much greater in the case of the long units, which required more than one sitting. The greatest saving in the short units was about 11 per cent. in the 20-line and the 50-line units. In the larger units the saving was about 20 per cent in one 60-line test and 22 per cent. in the other, 17 per cent. in the 120-line test and 20 per cent. in the 240-line test" (76, p. 138). Lakenan's work (50), carried out with prose material, verified Pyle and Snyder's findings on the length of material and gave evidence that subjects who at first learned or retained better by the part method, after additional practice, gave better results by the whole method. Ephrussi (22), using recall with adult subjects, found the whole method superior to the part method except for nonsense syllables and vocabularies; but the later experiments of Neumann (61) and of Brown (9) give the superiority to the whole method when vocabulary material is used.

We come now to the consideration of the work which is less favorable to the whole method. Winch, working with 38 school children, organized two groups of subjects of equal age and of equal ability as judged by their performances in preliminary learning tests. One group learned poems by the part method, and the other, by the whole method. In the former case the subject was instructed to learn a line at a time until one stanza was mastered. After the first stanza was learned, the second was begun. The first two stanzas were than repeated until they could be recalled correctly, and a new unit was then learned. Sample results are shown in Table 4, and Winch's comments upon the work are as follows: "There is a differential advantage of 26 per cent. in the four final tests in favor of the part method: this differential advantage falls to 5½ per cent. in the case of *Dora* [one of the poems] and rises to 52 per cent. for the *Seagull*" (118, p. 70). In this and other experiments by Winch, the part method tends to be superior when the material to be learned is short, or markedly disjointed and rhymed.

Reed (78), also working with poetry and using 113 college students, states that the whole method was superior for only 26 students. Thirty-one students did best with the pure part method in which stanzas were learned separately and then as a whole. Fifty-six students did best with

TABLE 4
Whole vs. Part Methods in Learning Poetry
(From Winch)

N	Age	Total marks in preliminary tests	Total marks in exper.	Marks for *Dora*	Marks for *Seagull*
			Results by the Part Method		
19	11 yrs. 8 mos.	273	279	173	125
			Results by the Whole Method		
19	11 yrs. 7 mos.	273	222	164	82

a progressive part method in which stanza 1 was learned, then stanza, 2, then 1 and 2, then 3, then the first three together, etc.

Pechstein (67) compared maze learning by the whole method with that by several varieties of the part method, using as subjects both white rats and college students. The maze was constructed of four well-defined parts which could be learned independently one of the other. The part methods of training adopted were as follows: (1) Pure part. (2) Progressive part. (3) Direct repetitive, where part 1 was first learned, then the subject ran through this part and on into part 2.

TABLE 5
Whole vs. Part Learning for Rats and Human Subjects
(From Pechstein)

Method	No. of rats	Av. trials	Av. time sec.	Av. total errors
Progressive part	9	11	662	65
Reversed repetitive	8	17	882	76
Direct repetitive	11	21	1442	142
Whole—returns prevented	9	30	1666	111
Pure part	9	30	1907	199
Whole—returns allowed	12	27	4174	217

Method	No. of human subjects	Av. trials	Av. time sec.	Av. total errors
Progressive part	6	10	352	57
Direct repetitive	6	11	618	96
Whole—returns allowed	6	12	641	126
Whole—returns prevented	6	17	541	81
Reversed repetitive	6	22	1014	226
Pure part	6	23	1220	237

When this was mastered, part 3 was added to parts 1 and 2, after which part 4 was added in the same manner. (4) Reversed repetitive. In some of the experiments by the whole method the subject was allowed to retrace as much as he would. In other experiments the amount of return permitted was no greater in the whole method than in the part method. Table 5 summarizes some of the results secured. The number of subjects in each group is too small to make the averages dependable, and the records undoubtedly have a low reliability. As the results stand, however, the progressive part and the direct repetitive part methods are superior to the whole method for both groups of subjects. For the rats the whole method with returns permitted, which is the usual method followed in maze experimentation, is the least efficient.

Our brief analysis of the problem of whole versus part learning can only lead to the conclusion that in some way as yet undetermined learning and retention are influenced by the whole versus the partial presentation of the stimulus series at each trial. Which method will prove most advantageous will probably vary with the character of the stimulus material, with the individual subject experimented upon, with the subject's age, with the amount of training which he has had, and with other as yet undetermined factors. We cannot generalize and say without qualification that either method is superior to the other.

The Distribution of Effort

The problem of the distribution of effort concerns the influence upon learning and retention of the introduction of rest periods between each two successive periods of practice. In the study of this problem we may vary the length of the periods of rest while keeping the magnitude of the work periods constant, or the conditions may be reversed. In the first instance the period of work may consist of one trial in the maze and the rest periods may be of one, two, and three days. One group of subjects would work for one trial per day; one, for one trial every two days; and one, for one trial every three days. In the second instance one group of subjects would have one trial per day; another group, two trials per day; and a third group, three trials per day.

Many investigators, beginning with Ebbinghaus, have contributed to the solution of this problem. The experimental findings indicate that, within limits, learning is accomplished with less work when practice is distributed than when it is concentrated. This principle is drawn from experiments upon a great variety of problems: maze habits, problem boxes, verbal material, archery, and arithmetic. Ebbinghaus used nonsense syllables as the learning material. His results, secured with himself as subject, are as follows: Six lists of 12 nonsense syllables each were learned at a given time with an average of 410 repetitions. Twenty-four hours later the relearning of the lists required an average of 41 repetitions. One list therefore required 68.3 repetitions in immediate succession for initial learning and 6.8 repetitions for relearning. When on the other hand a list of 12 syllables was learned to the point of one correct recall on each of three successive days, the total figures for the three days of distributed effort are 17.5 trials, 12.1 trials, and 8.3 trials, or a total of 37.9. Twenty-four hours later, the 12 syllables were relearned, on the average, in 6.2 trials. Thus 37.9 repetitions distributed over three days gave essentially the same retention after 24 hours as did 68.3 repetitions in immediate succession.

Jost (39) made a more thorough study of the same problem. He also found that concentrated repetitions of the responses to be mastered gave a poorer retention than did the same number of repetitions distributed over a number of days. This was true when retention was tested by the relearning method of Ebbinghaus and also when tested by the method of reinstatement. Jost was also able to show that the explanation of the greater effectiveness of the distributed work did not lie in the lesser fatigue that might accompany the shorter periods of work.

The greater effectiveness of distributed work can also be shown in the curves of learning. Starch (90) had 42 subjects divided into four groups work on a problem of digit-letter substitution. "The first group worked ten minutes at a time twice a day for six days. The second group worked for twenty minutes at a time once a day for six days. The third group worked forty minutes at a time every other day for six days. And the fourth group did the entire task at one continuous sitting. In each case the total time was 120 minutes" (90, p. 211). Figure 10 shows the degree of skill in the substitution for successive five-minute intervals of work. There is very little difference between the 10-minute and the 20-minute groups, but the 40-minute group is decidedly inferior, and the group that did all of its work at one sitting made the worst record of all.

Two important studies of the effect of distributed effort on the learning of infrahuman animals have been made by Ulrich (107) and Warden (109). Ulrich tested white rats with one, three, and five trials daily on a problem in which they had to lift a latch in order to enter a box and secure food. Tests were also made with a maze and with an inclined-plane box. From the distribution tables which record both the number of trials and the number of days required to perfect the habits under the several conditions, the general conclusion can be drawn that the less frequent the trials the more economical the learning when measured by total trials, but the less economical when measured by the number of days required. Figure 11 presents one of Ulrich's distribution curves.

Warden's work was conducted on a maze. Table 6 gives the average number of trials required by the various groups to reach a degree of mastery where nine trials out of ten were without error. The table indicates that four groups were trained with one trial per day where the length of the interpolated interval varied from six hours with one group to three days with another group. The results suggest that the interpolation of a 12-hour rest period is the most efficient, whether it occur after 1, 3, or 5 trials. Furthermore, one trial is preferable to a sequence of three or five trials under all conditions here represented. Warden's experiment indicates clearly the method of determining the optimal combina-

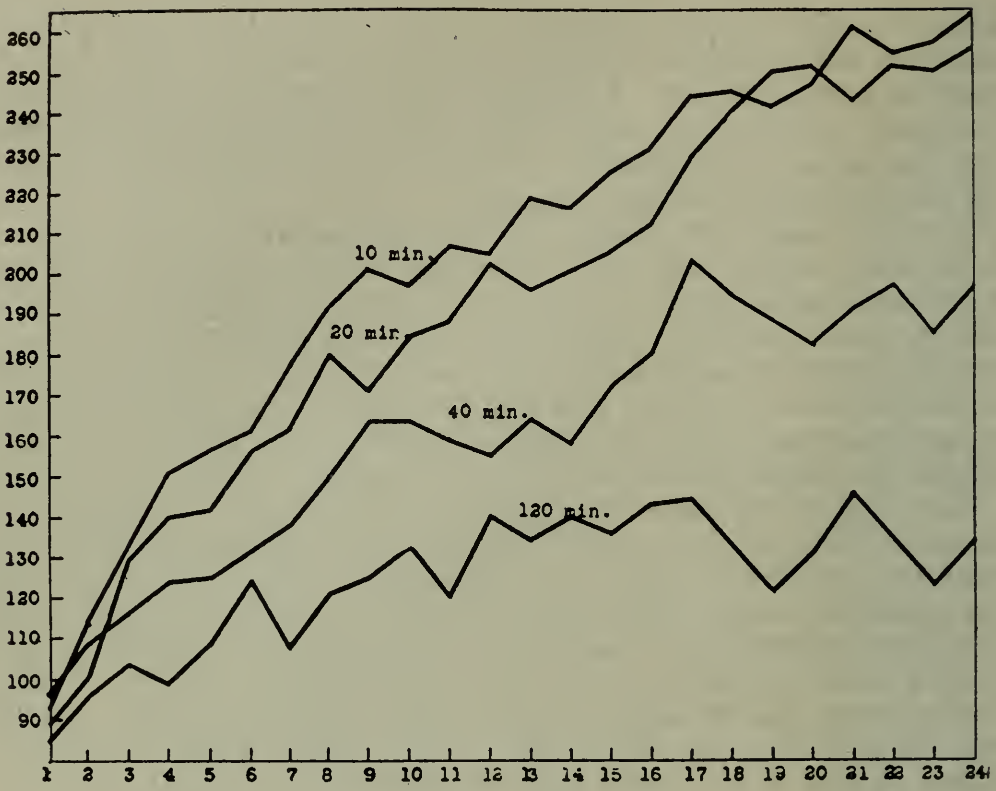


FIGURE 10

THE EFFECT OF DIFFERENT DISTRIBUTIONS OF EFFORT ON THE LEARNING CURVE The ordinate values represent achievement; the abscissa values, successive five-minute periods of work. (From Starch, *J. Educ. Psychol.*, 1912, **3**.)

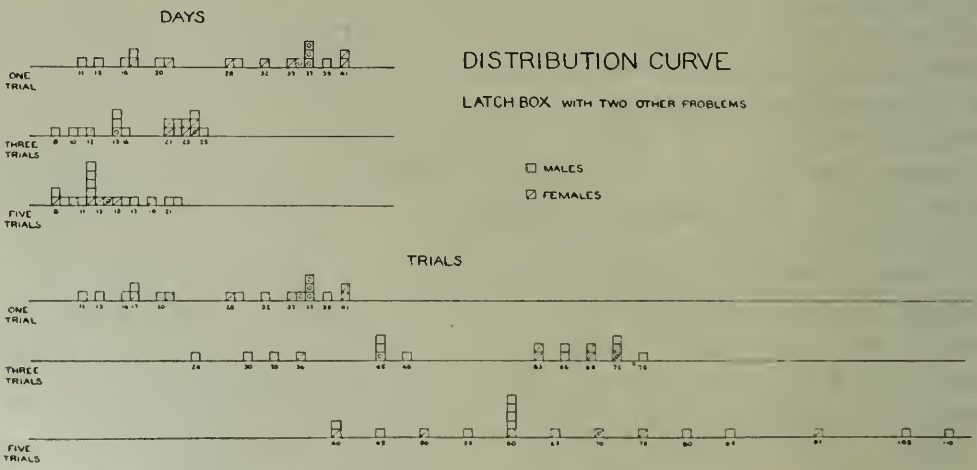


FIGURE 11

THE EFFECT OF THE DISTRIBUTION UPON THE NUMBER OF DAYS AND THE NUMBER OF TRIALS REQUIRED BY RATS TO MASTER A PROBLEM BOX (From Ulrich, *Behav. Monog.*, 1915, **2**.)

TABLE 6

EFFECT OF LENGTH AND FREQUENCY OF INTERVAL UPON MAZE LEARNING STATED IN TERMS OF THE MEAN NUMBER OF TRIALS REQUIRED FOR MASTERY

(From Warden)

Length of interval	Frequency of interpolated interval					
	After 1 trial		After three trials		After 5 trials	
	Trials	M. D.	Trials	M. D.	Trials	M. D.
6 hrs.	45.	12.	62.4	11.1	74.1	17.5
12 hrs.	36.8	9.7	54.8	12.1	63.6	11.4
1 day	46.4	13.1	65.4	18.4	86.1	15.2
3 days	71.9	13.5	91.1	24.2	86.1	18.4
5 days			129.7	26.	107.1	20.7

tion of length of interval and amount of practice for the acquisition of any stimulus-response coordination. However, the size of the M.D.'s in Table 6 and the low reliability of the maze make us cautious in accepting the data of this table at their face value.

Carr (14) in work with a stylus maze and Cummins (18) in experiments with a French-English vocabulary have shown that the effect of the distribution of effort varies with the point in the learning process at which the distribution is introduced. Carr finds that rest periods inserted in the first trials are more effective in reducing the total error score than if introduced in the later trials. Cummins' work indicates the desirability of short rest periods during the initial stages of learning and of longer intervals in the final stages. [See Ruch (83) for a general survey of the problem of distributed effort.]

Various theories have been proposed to account for the increased efficiency in learning and retention that comes with the distribution of effort. Müller and Pilzecker's doctrine of perseveration, which they elaborated in connection with the effect of interpolated work upon the retention of previously learned coordinations, has been applied also as an explanation for the present phenomenon. According to this hypothesis, the neural processes involved in behavior persevere for an appreciable interval of time after the behavior ceases and so are better retained if allowed to set without interference. This is a purely hypothetical explanation and one, moreover, which it is difficult to reconcile with the apparent fact that efficiency varies both with the length of the interpolated rest interval and with the quantity of work done during the work intervals. Lashley (52, 53) has suggested that there may be no general and fundamental law at the basis of the increased efficiency, but that specific factors, varying with the specific responses studied, may account for the results. As a concrete example he has been able to show, with rats learning the maze, that concentrated trials make for less diversity in behavior than do distributed trials, i.e., that with concentrated trials the same errors tend to be repeated to a higher degree than with distributed trials. In addition Lashley has the following comment to make: "In other forms of learning there are many agents such as fatigue and loss of interest in long practice periods which may interfere with efficient performance and so prolong the apparent learning time, while with verbal habits the possibility of practice outside of the experimental practice periods has not been altogether eliminated where short practice periods were used" (53, p. 366).

Age and Learning

As the human individual passes from infancy through adolescence and maturity to senescence, his apparent ability to cope with environmental difficulties changes. Our present problem concerns the possible influence of the age factor itself upon these changes. If it can be shown by careful experimentation that learning performance does vary with age, the phenomenon may be due either to intrinsic factors of neural growth and decline operative during the organism's life-history or to such extrinsic factors as habits which facilitate or impede learning and socio-economic conditions whose variations during the individual's lifetime may cause significant changes in the motivating processes involved in learning.

Our first problem is one of fact. Let us therefore inspect briefly some of the more important experimental studies of the problem. Hubbert (31), Liu (54), and Stone (93) have studied maze learning with white rats belonging to various age groups. Hubbert used the Watson circular maze with a *camera lucida* attachment (Figure 12), which made possible the accurate recording of the pathways followed by the subject. Table 7 summarizes her findings. Hubbert says: "Young rats learn the maze more rapidly than the old ones, the rapidity with which the habit may be formed decreasing with increase in age" (31, p. 54). An inspection of the table of results, however, reveals a large overlapping of the records made by the various age groups, which are also small in number of subjects; and Paterson (64) has pointed out the statistical inadequacy of the data to support such a conclusion. Hubbert also says that the most rapid improvement comes earlier in the learning process for younger than for older animals. This is not supported by Stone's later and more careful work.

Liu's investigation was conducted with rats of 30, 45, 60, 75, 100, 150, and 250 days of age. Table 8 shows the mean trial, error, and time scores for the different groups in learning one of Carr's mazes. It would appear from the data of this table that ability to master the maze increases with the age of the rats from 30 to 75 days and decreases thereafter. The results are much more definite than Hubbert's, and yet they are not convincing. It is impossible to tell whether they are the result of the accidental selection of animals, of the unreliability of the maze, or of a faulty control of incentives in the various age groups. We shall present below a quotation from Stone bearing upon the significance of this latter factor. Let us turn then to a consideration of Stone's experimental findings.

Stone used three different problems: an inclined-plane box, a modified

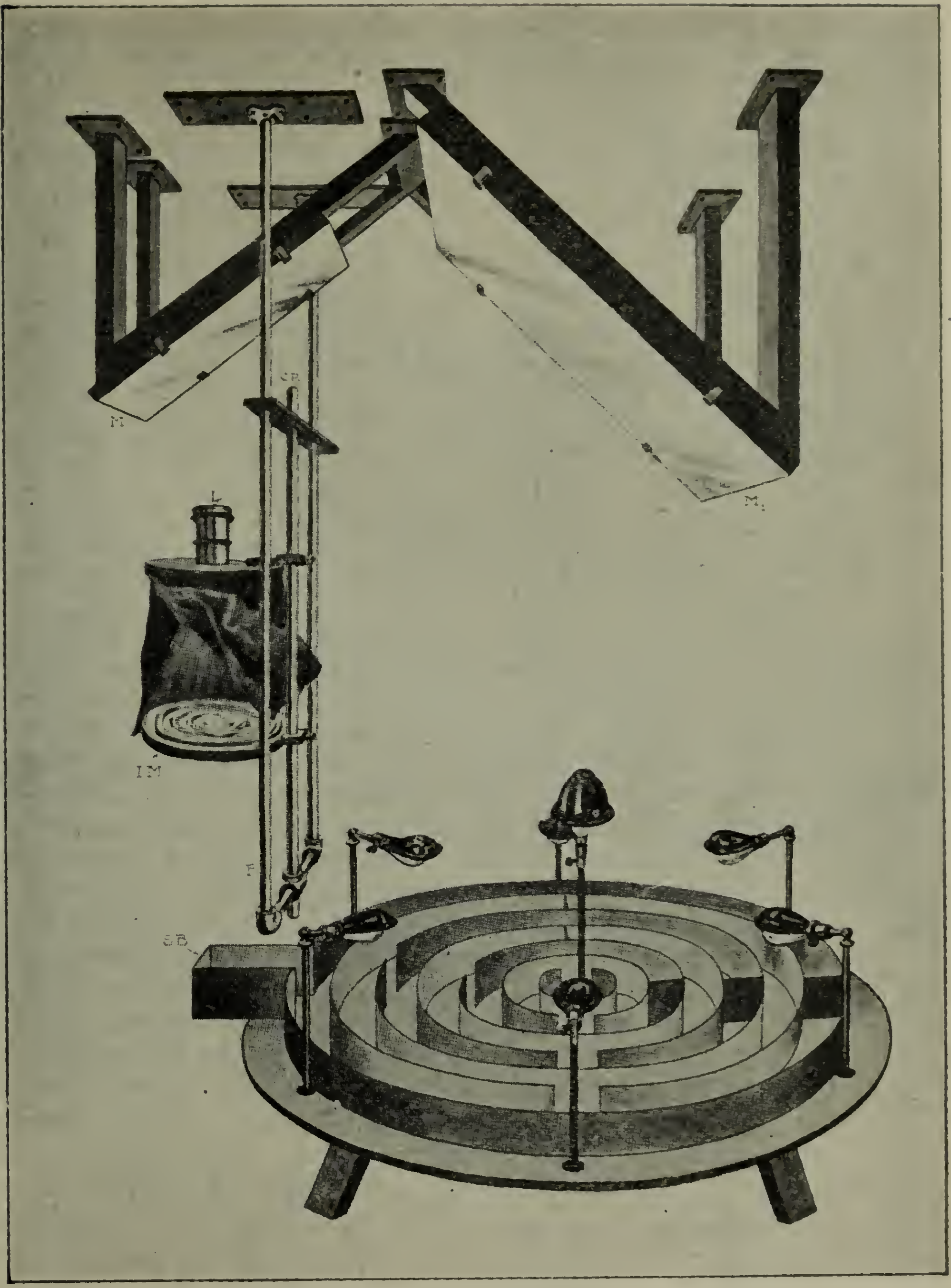


FIGURE 12
THE WATSON CIRCULAR MAZE WITH *Camera Lucida* ATTACHMENT
(From Watson, *J. Animal Behav.*, 1914, **4**.)

TABLE 7

Averages Required for Mastery of Maze

(Based on Hubbert)

Age	Trials	Time in min.	Distance in meters
25 days			
11 males	32	21.2	277
16 females	29	23.1	267
all	30	22.4	271
Range of score	14-51	6.4-64	139-418
65 days			
16 males	26.8	19.2	228
11 females	36.5	25.7	307
all	30.7	21.8	260
Range of score	14-65	6.4-73.1	91-750
200 days			
15 males	39.4	25.2	296
13 females	44.5	46.5	388
all	41.7	35.1	339
Range of score	14-112	8.2-94.9	125-912
300 days			
13 males	40.4	73.3	344
15 females	41.3	75.2	380
all	40.7	74.3	367
Range of score	14-78	15.9-227.3	117-723

Carr maze, and a multiple T-maze. The last problem gives the most dependable results, and it is from this one that we shall draw illustrative data. Rats of four different life-periods were used: 31 days, about halfway from infancy to puberty; 56 days, about mid-puberty; 456 days, about mid-life; and 730 days, somewhere between mid-life and senescence. Figure 13 shows typical results in terms of the elimination of errors during 30 trials. These results were all secured from rats whose diet was carefully controlled through its effect upon the rats' body weights. All age groups worked with as high a degree of hunger incentive as dietary restrictions could give. From the figure it is clear that no difference in the progress of mastery can be correlated with the different ages of the sub-

TABLE 8

A Summary of Liu's Data on Age and Learning in the Rat

Age	No. of animals	Av. trials	Av. total time	Av. total errors
30 days	40	53±2.35	1182±50.31	170±7.01
45 "	24	48±3.08	1155±71.66	151±8.12
60 "	24	39±2.94	1054±73.96	145±9.79
75 "	43	25±1.34	664±30.13	69±2.46
100 "	25	29±2.28	859±58.	83±5.
150 "	25	31±1.88	976±72.8	85±5.
250 "	25	40±5.37	1542±89.6	137±10.

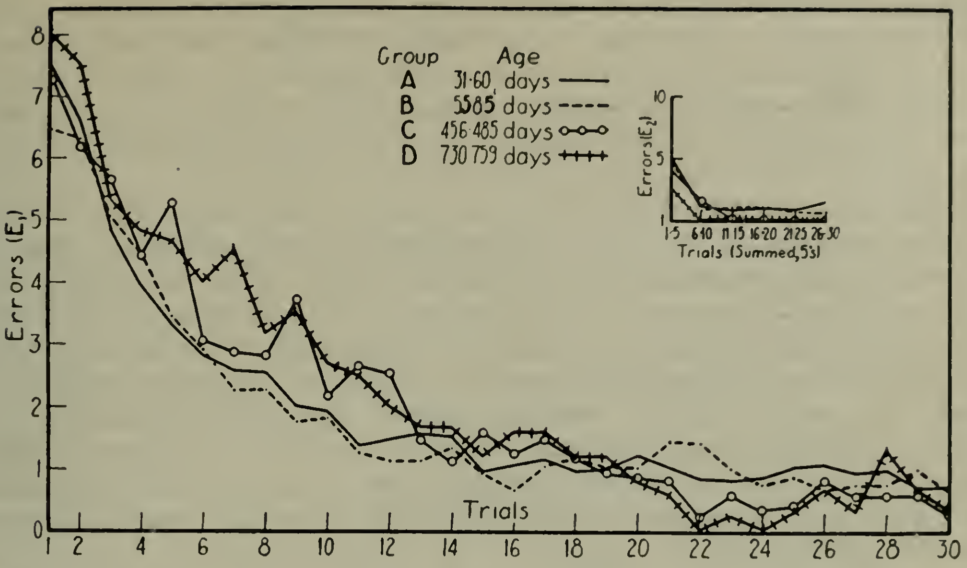


FIGURE 13

The Elimination of Errors in a Maze by Rats of Four Different Age Groups
(From Stone, *Genet. Psychol. Monog.*, 1929, **5**.)

jects. (Curves on the Vincent-Kjerstad plan would be more desirable from the standpoint of group comparisons. However, as we shall soon see, Stone's data in general fail to show any clear effects of an age factor, and Vincent curves which I have constructed from data supplied by Stone also show no age effects.) The number of trials required for mastery likewise fail to show significant age differences. The following statement by Stone is so important for the study of such differences, and also for other problems involving the maze, that I quote it at some length: "Striking would have been the *apparent age* difference obtained, if in a single experiment we had motivated young animals [strongly and the middle-aged groups less strongly]. In that case, the obtained results would have shown statistically significant inferiority in error scores for the two middle-aged groups and even would have yielded slightly inferior results for the older of the two middle-aged groups. With the prevalent rule of thumb and non-uniform methods of motivating animals by different experimenters working in the field of maze studies, the foregoing hypothetical case is not at all far-fetched. *In fact, it is exactly the result that an experimenter using present-day techniques is most likely to obtain, because, in the absence of adequate growth norms for the diet used, it is much easier to overlook the fact that young animals with constant weight increase while on the maze experiment may actually be subjected to a greater degree of chronic inanition than mature animals which are held at maintenance or allowed to increase or decrease slightly in weight.* . . . Age differences in maze records obtained under any condition in which either group is not exercising its maximum learning ability may be regarded with suspicion of being spurious, whatever their direction, until a scientific method of equating the motivation of different groups has been devised. Likewise, it is equally important that similarities in maze performances of rats not held to the maximum of their ability be regarded with suspicion of being spurious unless evidence of a less equivocal nature may be brought to bear upon the case in question" (93, pp. 103-105).

With the failure to demonstrate variations in learning ability with age in rats, we are left without a factual basis for the hypothesis that the plasticity of the nervous system varies with age. With the rats we are able to control the amount and character of the habits established prior to the experimental study in a way that is impossible with human subjects. We are thus with animal subjects in a highly advantageous position for discovering any intrinsic change in neural plasticity which might be present. With human subjects, when differences in the learning performance of individuals of different ages are found, it is impossible to separate the influence of previously established habits from the possible influence of a neural plasticity.

Let us consider briefly certain of the findings where human subjects have been used. In the section on Reminiscence we shall see some evidence that the recall of imperfectly learned response is partly a function of age. Thorndike and his colleagues (103) have made the most extensive study of different adult age groups, and numerous investigators have confirmed the popular impression that the ability to form habits improves during childhood. Thorndike presents Figure 14 as a fair generalization of the relation of age to learning in human subjects. His own data concern adults whom he has tested with a great variety of learning problems such as substitution tests, writing with the wrong hand, learning to typewrite, learning Esperanto, mastery of school subjects, and learning to respond to four signals with four simple acts. In this latter experiment 12 subjects ranging in age from 20-24 years were compared with 12 sub-

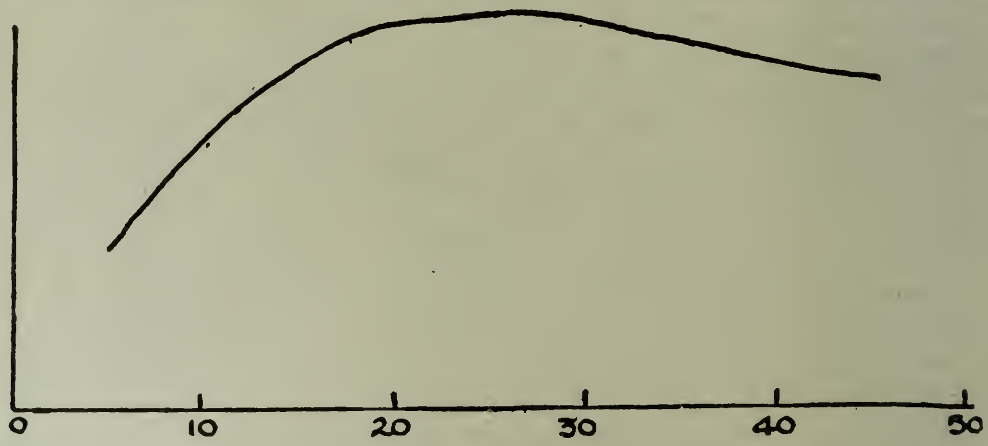


FIGURE 14
"The General Form of the Curve of Ability to Learn in Relation to Age" in Human Subjects
(From Thorndike, E L. Adult learning. New York: Macmillan, 1928. Reproduced by permission.)

jects over 35 years of age (average age, 42), the two groups having about equal ability on the CAVD behavior test. (See reference 102 for a description of this test.) The subjects were blindfolded and instructed to draw a line 3, 4, 5, or 6 inches long. Six hundred such lines were drawn by the subject while ignorant of the accuracy of his performance. Then for seven days, the subject drew 600 lines per day and was told each time whether the response was right or wrong. After this came a third test in which the subjects drew 600 lines without being told whether the lines were right or wrong. In the first period there was no essential difference between the age groups. In the last part of the seven days' work, however, the old scored only 83% as much as the young, and in the third test the old were only 70% of the young. Although the number of subjects is small, the results are in harmony with those derived from Thorndike's other work. The results are chiefly significant because the kind and amount of previous specific training possessed by the various subjects can hardly have influenced the relative performance of the two groups. Differences in motivation and in the adjustments of the two groups to the social aspects of the experimental situation, however, remain as possible explanatory factors.

Table 9 from Thorndike summarizes the results of his experiments. Thorndike comments upon this table as follows: "The facts of table [9] in general are fairly consistent with the hypothesis that the old are considerably inferior to those around 22 in general basic modifiability but compensate for this inferiority somehow (probably by better appreciation and organization and use of what is learned, possibly by greater interest)

TABLE 9

SUMMARY OF THE RELATION OF AGE TO ABILITY TO LEARN
(From Thorndike)

U, university group; P, prisoners; H, evening high school students; S, students in secretarial schools

Group	Ability	Gain in score			Per cent which O is of Y
		Y 20-24 yrs.	25-34 yrs.	O 35 yrs. & over	
U	Drawing lines				64
	Wrong hand writing	57	51	41	72
	Substitution (words)	28.0	27.9	22.8	81
	Learning code	10.4	8.3	6.3	61
	Esperanto	31.5	26.3	24.7	79
	Learning nos. to fit nonsense syllables	21.9	18.2	14.0	64
	University studies				over 100
P	Substitution (digits)	9.0	9.2	9.4	104
	Elementary school subjects	100	100	88	88
	Addition practice	5.0	4.6	4.8	96
H	Algebra, Civics, English, etc.				87 for 30 or over
S	Typewriting				95 for 30 or over
	Learning stenographic symbols				100 for 30 or over

when learning typewriting or stenography or school studies." "The general tendency from all our experiments is for an inferiority of about 15 per cent as a result of 20 years from twenty-two on. Learning representing an approximation to sheer modifiability unaided by past learning [the line-drawing experiment] shows considerable more inferiority than this. Actual learning of such things as adults commonly have to learn shows considerably less" (103, pp. 104 and 106).

Two different stories of the relationship of age and learning are told by the animal experimentation by Stone and the human tests by Thorndike. Even though practical learning situations give the curve of Figure 14 for human ability, there are too many factors involved other than those of neural development and regression to enable us to infer that these latter factors exist and are responsible for the shape of the curve. Thorndike distinctly recognizes such possibilities as the awkwardness of adults in certain social situations, the difficulty of securing the adult's complete cooperation, and the inhibiting effects of habits and customs. One cannot argue directly from rats to men, but until experiments upon man have been carried out with an equating of incentive and previous relevant training for the various age groups, we must remain skeptical of the rôle of intrinsic neural changes as a causal factor in those learning problems which the various age groups can all undertake.

THE RELATION BETWEEN SPEED OF LEARNING AND RETENTION

The experimental investigations of this problem indicate not only that the responses first learned are best retained but also that the individuals who learn most rapidly also tend to make the best records in retention tests.

Woodworth (120) conducted an experiment using 20 words of an Italian-English vocabulary as the stimulus material. After the subject had read the list once, the experimenter gave the Italian words as stimuli and the subject was to respond with the English equivalent. From 3-5 sec. were allowed for each response, and the subject was prompted and corrected when necessary. This training was continued until each correct response was given once. As soon as an Italian-English combination was learned to the point of one correct recall that pair of words was dropped from the list (to prevent overlearning), and the training was continued with the remaining pairs. When all had been learned, the entire list was again presented once. After an interval of 2-20 hours the Italian words were presented singly and the recall score for the corresponding English words was determined. The results were as follows:

Of the pairs learned in *x* readings *x*	*y* per cent were retained *y*
1	73
2	72
3	63
4	58
5	38
6-11	27

A part of this decrease in retention with increased difficulty of learning might be due to differences in the associative responses called out by the various words. Woodworth, however, reports that when only words not so supplemented are considered the relationship between speed of learning and excellence of retention remains the same.

Important work on the retention of learned material by individuals who learn quickly and by those who do not has been carried out particularly by Norsworthy (63), Pyle (74), and Lyon (56). Norsworthy, using 83 subjects and a German-English vocabulary of 1200 words, found that the quicker the learning, the better the retention. In this experiment the amount of time during which the subjects worked was kept constant, and the speed of learning was shown by the number of correct responses which the subjects learned. A high positive correlation (0.41—0.50) was found between the number of responses learned and the number that could be correctly reinstated. It was also found that the subjects who learned over 700 word responses (the quick learners) had a better relative retention score than did those (slow) learners who mastered only 33 word responses.

Lyon's work was carried out with three methods: In method 1 the subject learned the required responses and, after an interval of delay, attempted to reinstate as many as possible without having had the original material represented. Method 2 immediately followed method 1. The stimulus material—words, digits, nonsense syllables, or prose—was again presented once, and the subject was then asked to reinstate the appropriate responses. Method 3 immediately followed method 2. In this case the stimulus material was again presented to the subject, and a complete relearning was required. The chief defect in Lyon's method seems to lie in his lack of control of the learning periods. In place of rigidly controlling the amount of time and the number of trials required for mastery, he permitted the subjects to study until they were certain that they could reinstate the required responses without error. This method certainly resulted in a variable and undetermined amount of overlearning. Table 10 shows results based upon the records of 24 normal-school seniors who were divided into two groups (upper and lower halves) on the basis of the quickness of initial learning.

An examination of this table shows that the quicker half of the learners was superior in retention to the slower half with all materials when method 1 was used. With method 2 the slower learners surpassed the quicker learners in the case of digits only. With method 3 the slower learners retained more in the case of digits, words, and poetry than did the quicker learners. Lyon also calculated coefficients of correlation between learning and relearning. Although clear-cut relationships were not generally demonstrated, the results favor the conclusion that quick learning is accompanied by good retention.

Numerous studies have been made of the relation between the amount learned, as measured by a test of recall immediately after learning, and the amount retained at some later time. These studies have shown a positive correlation between immediate and delayed recall which may be as

TABLE 10

Relation of Speed of Learning to Retention

(From Lyon)

Materials	Subjects	Method 1	Method 2	Method 3		
		% reinstated	% reinstated	Av. learning time	Av. relearning time	% saved
Digits						
	quicker half	27	34	5.5	2.0	63
	slower half	22	36	15.0	4.0	71
Nonsense syllables						
	quicker half	26	44	24.1	6.5	71
	slower half	21	36	38.0	12.1	68
Words						
	quicker half	23	44	10.0	5.5	40
	slower half	20	35	14.8	5.7	61
Prose						
	quicker half	59	80	14.3	4.8	68
	slower half	46	67	27.3	9.1	66
Poetry						
	quicker half	59	73	7.0	2.8	61
	slower half	40	57	14.5	5.2	65

high as 0.82 found by Winch (118), or 0.71, found by Gordon (26). The score for immediate retention, however, can represent speed of learning only where the learning occupies a constant period. This condition is fulfilled in Gordon's study, and the coefficient of 0.71 may be regarded as an indication of the relationship between speed of learning and retention. Brown (8), also working with a constant time period for learning, finds positive correlations between speed of learning and retention. Brown believes "that a positive relation is normal and that when conditions are favorable the correlation is very close between amount learned and amount retained." This point of Brown's is further brought out by the fact that the methods of stimulus presentation, e.g., the distribution of stimuli as opposed to their concentration, which favor learning also favor retention.

The Influence of Active and Passive Recall upon Retention

In the ordinary experiment upon the learning of verbal habits, the subject is instructed not to practice on the problem between trials, i.e., not to attempt to reinstate the verbal responses except when the stimuli are presented by the experimenter. Obviously such *active* recall will effect the rate of learning. If erroneous responses are made, they will tend to retard learning; and if correct responses are made, they will facilitate learning. In any case the experimenter would not have a record of what responses had been made and so would not be able adequately to interpret the results.

The topic of our present discussion raises a slightly different aspect of the problem of repetitions not controlled by the experimenter: Does the *active* reinstatement of responses have a more advantageous effect upon learning than *passive* reinstatement? The terms *active* and *passive* are not aptly chosen to designate the experimental facts, but our descriptions of experimental procedures will show that active reinstatement occurs in the absence of the stimulus material and that passive reinstatement occurs in response to the presentation of that material. Reading and recitation are other terms used to designate the same phenomenon. The investigation of recall with and without the aid of the stimulus was begun in 1907 by Witasek (119) and continued particularly by Katzaroff (41), Kühn (49), Knors (44), Gates (25), and Skaggs (86). Although the experimental methods used by these investigators have varied greatly, their results justify the general conclusion that under certain conditions the reinstatement of a response, without the aid of its usual stimulus, aids learning more than a reinstatement called forth by the presentation of that stimulus.

Gates's work, the most complete recent investigation of the problem, was carried out upon large numbers of school children and a few adults. The material to be learned was either nonsense syllables or consecutive verbal material patterned after the biographical sketches in *Who's Who*. Table 11 shows the various combinations of active and passive recall (recitation and reading) utilized with school grades 4, 6, and 8. In method 1 the subjects read the material for 9 minutes. In method 2 the reading occupied the first 7 mins. and 12 sec. of the study period, and the final 1 min. and 48 sec. was devoted to recitation. When the subject recited, he did so quietly and to himself. He was instructed to glance at the material to be learned only when he was otherwise unable to reinstate the appropriate responses. The remaining methods of the table are to be similarly understood. At the close of the 9 minutes of work by each method, the subject was asked to write down as much of the material as possible. These records were then scored by a conventional method.

TABLE 11

METHODS USED BY GATES IN THE STUDY OF READING AND RECITATION

Method	Time of reading	Time of recitation	% Reading	% Recitation
1	9 min.	0	100	0
2	7 min. 12 sec.	1 min. 48 sec.	80	20
3	5 min. 24 sec.	3 min. 36 sec.	60	40
4	3 min. 36 sec.	5 min. 24 sec.	40	60
5	1 min. 48 sec.	7 min. 12 sec.	20	80
	With biographical sketches method 6 was added			
6	54 sec.	8 min. 6 sec.	10	90

Table 12 is based upon nonsense material and shows the average scores and their P.E.'s for each of the five methods. These results indicate that as the amount of recitation increases, within the limits tested, the average recall also increases. Table 13 is based upon biographical material. The results indicate that for all four grades the efficiency of recall increases up to a certain point, with the increase in the amount of recitation, and declines after that. Of the six methods employed no. 5 is the most advantageous for grades 6 and 8, and no. 4, for grades 4 and 5. Where an interval of from three to four hours elapsed before the subjects were required to write down the material, similar results were secured; but there was a greater difference between the worst and the best methods than was found for immediate recall.

TABLE 12

Average Score for Each of Three Grades for the Various Methods Using Nonsense Material

(From Gates)

Method		1	2	3	4	5
Grade 8	Av.	16.92	23.86	25.79	27.28	35.51
	P.E.	0.61	0.69	0.65	0.66	0.86
Grade 6	Av.	13.21	20.18	22.64	25.15	30.52
	P.E.	0.61	0.84	0.60	0.91	1.07
Grade 4	Av.	9.45	12.00	16.10	16.95	20.03
	P.E.	0.57	0.46	0.56	0.75	0.79

TABLE 13

Average Score for Each of Four Grades for the Various Methods Using Biographical Material

(From Gates)

Method		1	2	3	4	5	6
Grade 8	Av.	20.77	22.39	24.84	24.95	25.28	23.75
	P.E.	0.72	0.87	0.70	0.69	0.50	0.82
Grade 6	AV.	15.13	16.55	18.01	17.70	17.77	16.63
	P.E.	0.75	0.59	0.69	0.68	0.82	0.68
Grade 5	Av.	11.79	13.95	15.21	15.96	15.33	15.74
	P.E.	0.40	0.43	0.48	0.56	0.50	0.55
Grade 4	Av.	14.61	16.91	16.36	18.81	17.62	17.20
	P.E.	0.77	0.78	0.86	0.77	0.70	0.71

Gates's results indicate that a larger and larger amount of recitation relative to reading when inserted earlier and earlier in the learning period significantly affects the efficiency of work. Various explanations have been offered for this general phenomenon. It seems probable that recitation as here conducted derives its advantage from the opportunity which it affords the subject for correcting his errors and from the fact that during recitation the subject is practicing the type of response which he will be required to make finally, viz., the recall of the learned material without dependence upon the text. This latter point is particularly well supported by the work of Katzaroff (41) who used the paired-associates method of recall both for the recitation and for the final test of retention.

In spite of the fact that careful laboratory investigations of this problem reveal that the proper introduction of recitation into the study period is advantageous to learning, the extension of these conclusions to the schoolroom is not justified. It may well be that a detailed and accurate scoring of returns from a carefully controlled laboratory experiment will reveal advantages of certain methods of work and yet all of these advantages may either be lost in the hurly-burly of schoolroom conditions or fail to be reflected in the rough scoring methods applied by teachers. Furthermore, the tables show that recitation aided the mastery of the biographical material less than it did the nonsense material. This fact we get by comparing the difference in score between the poorest and the best methods for each material. Thus for nonsense syllables the poorest method was no. 1 and the best was no. 5, which was 100% or more better than no. 1. In grade 8 for biographical material the best method is only about 20% better than the worst.

The Influence of Activity upon the Retention of Previously Established Habits

Let us assume that a subject has been trained for any given number of trials on a problem like repeating nonsense syllables or running a maze. During this training certain changes have been set up in his nervous system in virtue of which his behavior is different at the end of the training from what it was at the start. If training is now discontinued on the problem, as time passes, the subject shows less and less of the effects of his training, no matter what method of testing retention we use. Recall becomes impossible; the subject can no longer distinguish between the material on which he was trained and new material (the age of stimulus method); and even the relearning time shows no saving over that of the original learning. Theoretically such a loss of retention might be brought about merely as a result of the lapse of time, as a result of the metabolic processes in the body whereby the neural changes set up by training would gradually fade away. In order to prove that lapse of time itself brought a loss of retention, it would be necessary to conduct an experiment in which the influence of the activities interpolated between the original learning and the later test of retention could be ruled out as explanatory factors in the loss. The conventional method of conducting an experiment upon the influence of interpolated activities is as follows:

Experimental Group	Control Group
Training on habit no. 1	Training on habit no. 1
Rest, i.e., no activity	Interpolated work
Retest of habit no. 1	Retest of habit no. 1

It is impossible, however, in a living organism to have a period of "no activity." All that we can do experimentally is to vary the character and amount of the activity which is interpolated between the original learning and the retest period and to study the influence of such changes upon retention. We shall present data later showing how the loss of retention varies with the lapse of time; but as the lapse of time increases, the quantity of interpolated activity also increases. It is to this latter factor that we must look for the explanation of the loss. When investigators speak of periods of rest or of the mere lapse of time, we are to understand that the ordinary uncontrolled routine of the subject's life filled the period in question, or that some activity like newspaper reading (see experiments below) was present.

Ebbinghaus was the first experimenter to attack the problem of loss of retention as a result of increasing the quantity and duration of the interpolated activities. Müller and Pilzecker were the first to vary experimentally the character of the interpolated work and to study the resulting changes in retention. We shall begin our examination of the experimental evidence bearing upon the loss of retention with a brief survey of the work conceived in the manner of Müller and Pilzecker and then study the problem from the angle suggested by Ebbinghaus' work.

Müller and Pilzecker (60) used both the paired-associates and the relearning methods. The material to be learned was nonsense syllables presented on a revolving drum. The interpolated work was varied in a number of ways: (1) the subject rested for the entire period; (2) other lists of nonsense syllables were learned; (3) the subject studied pictures in order to describe them later; and (4) the material of nos. 2 and 3 sometimes came immediately after the completion of the original learning and sometimes after intervals of rest. Condition 1 was the standard with which the results of the other conditions were compared. Although very few subjects were used, the results indicate that when conditions 2, 3, and 4 follow learning, the scores on the retention tests are lower than when condition 1 follows learning. The use of nonsense syllables in the interpolated work was no more disadvantageous than the study of pictures. Retention scores are less if the interpolated work comes immediately after learning than if a rest interval intervenes before the interpolated work is done. (In order to test this phase of the problem, called the temporal position of the interpolation, the interpolated period might be 8 min. long divided into 4 min. work and 4 min. rest, 4 min. rest and 4 min. work, or all 8 min. either rest or work.) These findings of Müller and Pilzecker have very largely set the problems for later investigators, although the general fact of the disadvantageous effect of interpolated activities is the only result upon which most students are agreed.

Robinson (79) has made careful studies of the effect of various kinds of interpolated work and also of the influence of the temporal position of the work upon the retention scores. In both cases the results disagree with those of Müller and Pilzecker. Table 14 shows Robinson's results on the influence of temporal position. The interpolated periods (conditions) are each 20 min. long. In condition A, the 20 min. were spent in reading a newspaper. Conditions B, C, D, and E are described in the table. After eight exposures of a list of 10 three-place numbers, a period of 20 min. was interpolated after which the subject wrote down as many of the numbers as possible. Table 14 shows how the amount recalled and the errors made varied with the nature of the interpolated period. Condition A is notably better than B, C, D, or E, but there is no essential difference between these four conditions. Although Robinson's

TABLE 14

THE INFLUENCE OF THE TEMPORAL POSITION OF THE INTERPOLATED WORK UPON RETENTION SCORES

(After Robinson)

Condition	Interpolated period	Amount recalled	Errors of recall
A	20 min. reading	29.1	6.1%
B	5 min. study of 3-place nos. 15 min. reading	18.6	31.9%
C	5 min. reading 5 min. study of 3-place nos. 10 min. reading	19.7	29.0%
D	10 min. reading 5 min. study 5 min. reading	19.2	32.6%
E	15 min. reading 5 min. study	18.4	30.3%

work is undoubtedly dependable for the conditions under which he worked, the divergent results of other studies made under different conditions by Skaggs (87) and Whitely (114, 115) must make us cautious in generalizing Robinson's results.

Historically one of the most important problems concerning the influence of interpolated activity upon retention may be formulated as follows: How does retention vary as the similarity between the original learning and the interpolated work varies? Müller and Pilzecker found no difference in the disadvantageous effects of nonsense syllables and picture study upon the recall of nonsense material. This would indicate that the similarity of original and interpolated work is not an important factor in the problem. This result, however, is not confirmed by the later investigations of Robinson, Skaggs, and Whitely. Whitely, for example, examined the variation in the retention of words and phrases when just prior to attempts at learning and recall the subject was given either a résumé of material within the field of the original learning or a résumé from some field quite different from that of the material to be learned. While the difference was small, it indicates that recall was poorest when the résumé was drawn from the same field as the material to be learned. Robinson secured like results indicating the positive rôle of similarity when he used numbers for the original learning material and either numbers, nonsense material, poetry, multiplication, or prose reading for the interpolated work.

Skaggs and Robinson point out certain interesting theoretical aspects of the similarity problem. If the interpolated work is identical with

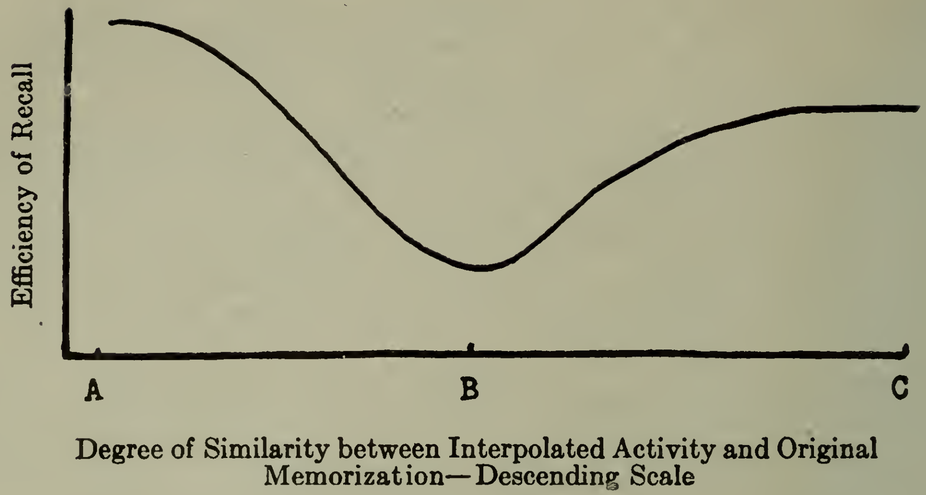


FIGURE 15

The Theoretical Relationship Between Efficiency of Recall and the Similarity between Interpolated Activity and Original Activity
(From Robinson, *Amer. J. Psychol.*, 1927, **39**.)

the original learning, this amounts to continued training and no loss of retention will occur, rather there will be an increase in efficiency due to the prolongation of practice. As this identity decreases with the introduction of a less and less similar material, retention of the original material becomes less and less until a minimum is reached. Beyond this point further decreases in similarity, i.e., increases in dissimilarity, have less effect upon retention, and the retention score thus rises. (See Figure 15.) Robinson (80) has sought to verify this theoretical relationship between original and interpolated work. The experiment was conducted by the memory-span method with consonants for the original and for the interpolated material. Where 12 consonants were used, the first six were considered the original learning material and the last six were considered the interpolated work. The recall score was the number of the first six consonants correctly reproduced. Dissimilarity was varied from the condition where all of the last six letters were different from the first six to that where they were identical with the first six. There might thus be from one to six letters common to the two sets. The results show that the score falls regularly from that with six letters in common to that with none in common, but the curve does not rise again as the theoretical consideration indicated that it should. Perhaps the difficulty lies in the concept of *similarity*. Robinson interprets degrees of similarity in terms of the identity of letters. Harden (27) used numbers and letters in a study of the same problem and found that the recall score again mounts when the interpolated material is numbers and the original material is letters. The future development of the problem of similarity of interpolated work depends upon the ability of experimenters to approach the problem quantitatively as Robinson and Harden have done. At present, experiments clearly indicate that one interpolated activity may differ markedly from another activity in its effect upon retention, but the reasons for the difference are as yet undetermined.

Heine (28) was unable to find disadvantageous results for interpolated work when retention was tested by the age of stimulus method. This was probably because the original material was so well learned that the interpolated activity used had but little effect. Since the age of stimulus method will reveal traces of retention when the recall method will not, the former method would fail to show a loss of retention unless that loss were considerable.

We come now to the consideration of experimental data which bear upon our theoretical analysis at the beginning of the present section. Is the loss in retention under conditions of *disuse* due to the fading away of neural changes or to the disrupting influence of subsequent activities? Let us first note the relationship between amount of material to be learned and loss of retention where the ordinary activities of the subject fill in the period between learning and the retest. Ebbinghaus (21, p. 84), working by the relearning method, found a saving of 33% after 24 hours for a list of 12 nonsense syllables. For 24 syllables the saving was 48%; and for 36 syllables, 58%. This general relationship between amount of material learned and retention after periods of no practice has been found by other investigators. Thus Robinson and Darrow (81) found that retention was greater as the amount of material to be learned increased. Their results show 60% recall for 4 numbers, 66.5% for 6 numbers, 66.6% for 8 numbers, and 70.8% for 10 numbers. Recall was tested after a 15-minute interval devoted to casual newspaper reading. If now the reading is replaced by a study of other lists of numbers, the degree of retention is further decreased, but the longer lists of numbers are less affected than the shorter lists. The general conclusion is indicated that the decrease in the retention of previously learned material which is brought about by interpolated activities varies inversely with the length of the material learned, which is exactly the effect that Ebbinghaus assigned to the influence of the lapse of time.

So far we have considered experiments where, after learning a given response, the subject is either allowed to continue his routine activities or is subjected to training of a controlled character. The former case is called the condition of rest in comparison with which one estimates the influence which the activities of the latter case have upon the retention of the learned response. We may now ask what effect the activities of the alleged rest period have upon retention by comparing the results for the rest period with those for a period of equal length where the subject sleeps, a condition of much decreased activity. Jenkins and Dallenbach (37) have performed such an experiment for periods of one, two, four, and eight hours. Series of 10 nonsense syllables were employed, and the two subjects used were tested by the method of recall. Great care was exercised to make the tests comparable for the two interpolated conditions of sleep and waking. Figure 16 shows the results. After one hour's sleep seven syllables were recalled, a loss of only about 28%. After one hour of routine working activity, the loss was about 46%. The noteworthy point in the figure is the relatively slight forgetting after sleep where the activities which might interfere with retention have been reduced to a minimum. [Nicolai, in his work (62) cited in the section on Reminiscence, also found retention better after sleep.] The authors also sought to rule out the influence of interpolated activities in another experiment where they utilized a hypnotized subject. The subject learned the syllables while hypnotized. He was then brought

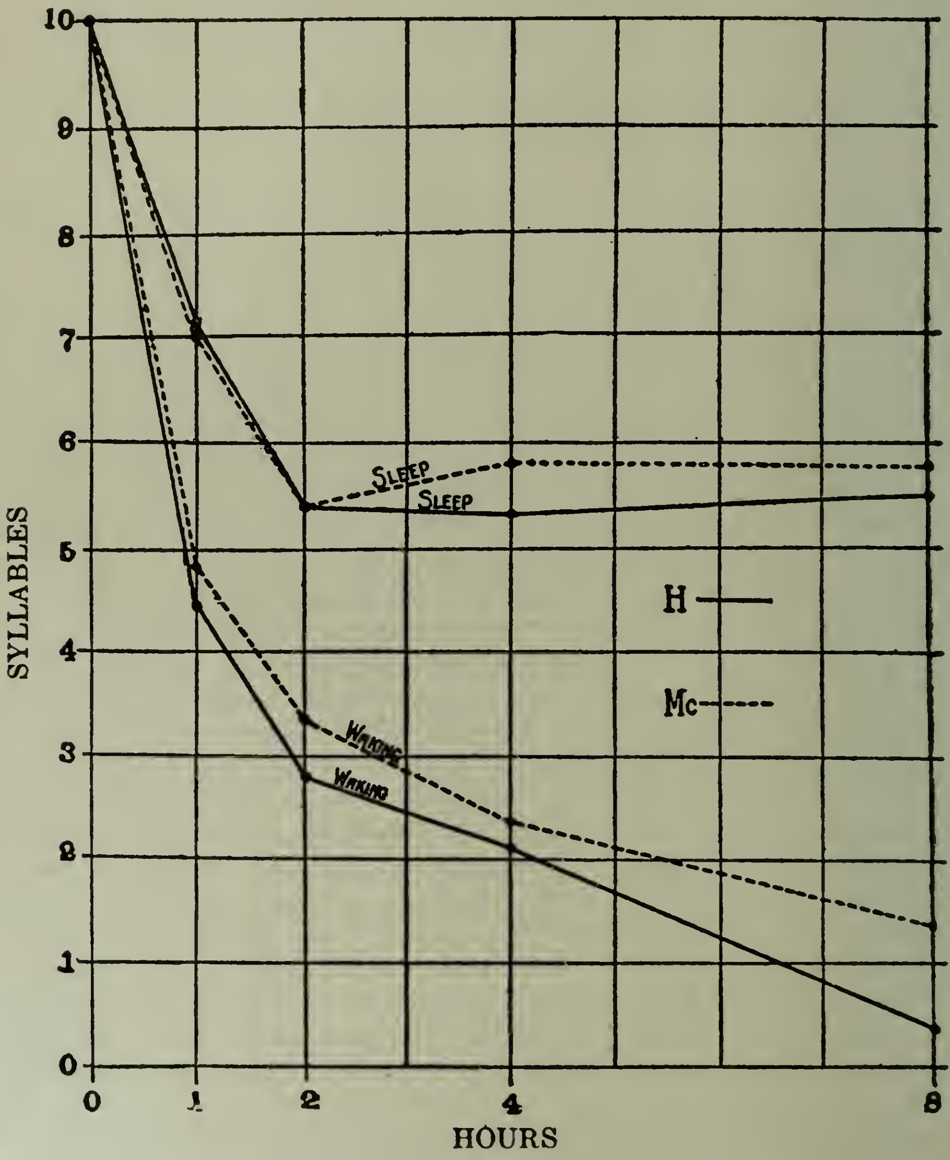


FIGURE 16

The Number of Syllables Recalled After Various Intervals of Time During which the Subject was Either Asleep or Awake

(From Jenkins and Dallenbach, *Amer. J. Psychol.*, 1924, **35**.)

out of the hypnotic state and allowed to continue his daily duties until the time for recall when he was again hypnotized for the test. Recall was perfect after intervals of two, four, and eight hours, and was 80% perfect after 24 hours. The position taken in the present section is in harmony with that of Jenkins and Dallenbach when they say: "The results of our study as a whole indicate that forgetting is not so much a matter of decay of old impressions and associations as it is a matter of interference, inhibition, or obliteration of the old by the new" (37, p. 612).

THE RATE OF FORGETTING

In the preceding section we have discussed certain aspects of our present problem. We saw in Figure 16 that the amount of loss of retention after different intervals of time depended upon the character of the activities which filled the intervals. These curves show a rapid initial loss, and this characteristic is also present in the data of Tables 15 based upon the findings of Ebbinghaus and Radossawljewitsch (77). The explanation of the rapid initial loss is not entirely clear, but the most reasonable explanation seems to lie in the activities of the organism which immediately succeed the learning period. This view is supported by the work of Jenkins and Dallenbach which we have already presented. It is probable that such experiments, involving interpolated intervals of sleep, would show no rapid initial drop if it were possible for the subject to go to sleep immediately after learning.

There is no one curve of forgetting; rather there are many curves the natures of which vary with such factors as the following: the degree of mastery of the initial learning, the character of the material originally learned, the speed of the initial learning, the amount and character of the activity which follows the learning, and the method by which retention is tested. In the present section we shall consider only the second factor, i.e., the relation between loss of retention and the character of the material originally learned.

Such conventional verbal material as prose, poetry, or even unrelated words are better retained than the relatively novel verbal material composed of nonsense syllables. Table 15 shows the difference between such materials when retention is tested by the savings method. Experiment has not made precise the reason for this difference between the two types of material, but theoretical analysis indicates pretty clearly that the advantage of the conventional material lies in the range of established associations which are present when the experiment begins. This insures not only a greater number of cues for the recall of the required responses but it also makes a large and uncertain degree of overlearning inevitable. And responses that are overlearned are better retained than those which are not so firmly established. [See, for example, Krueger (48)].

It has been widely held that verbal habits are less well retained than non-verbal habits. Watson (111, pp. 307-8), for example, writes: "It is thus seen that the rate of deterioration in habits such as we have considered [typewriting and ball-tossing], while in all cases positive, is very

TABLE 15

THE RATE OF FORGETTING FOR DIFFERENT MATERIALS
(From Radossawljewitsch and Ebbinghaus)

Period since learning was completed	Radossawljewitsch Percentage of saving, nonsense material	Percentage of saving, poetry	Ebbinghaus Percentage of saving, nonsense material
5 min.	98	100	—
20 min.	89	96	59
1 hr.	71	78	44
8 hrs.	47	58	36
24 hrs.	68	79	34
2 days	61	67	28
6 days	49	42	25
14 days	41	30	—
21 days	37	48	—
30 days	20	24	21
120 days	3	—	—

slow indeed. Ordinary observations show that the same is true in regard to swimming, skating, dancing, tennis playing, and skilled mechanical work. We shall see . . . . that this is in marked contrast to the rapid deterioration observed in habits which belong primarily to the language groups. There the deterioration is so rapid that in some cases, for example, in the learning of a series of nonsense syllables, the organization is lost in from fifteen minutes to half an hour so far as concerns the subject's ability to speak or write the words." Ordinary observations, however, are not adequate for the solution of the problem. Indeed it would be possible for any teacher of swimming or dancing to cite numerous cases where certain responses have appeared correctly once (which cor-

TABLE 16

RETENTION OF TYPEWRITING SKILL AS SHOWN BY NUMBER OF STROKES ON MACHINE PER 10-MINUTE TEST
(From Book)

Tests	1	2	3	4	5	6	7	8	9	10	Av.	Av. correct strokes	% error
Last reg. prac. Jan. 7-16, 1906	1503	1509	1404	1572	1494	1436	1501	1455	1508	1698	1508	1475	2.21
1st memory test June 1-10, 1906	1365	1421	1421	1433	1529	1443	1523	1504	1313	1472	1443	1391	3.54
2nd memory test, June 1-10, 1907	1390	1344	1345	1537	1681	1694	1634	1845	1761	1850	1611	1560	3.15

responds to one correct recall of a list of nonsense syllables) and yet are forgotten, i.e., cannot be reinstated, five minutes later. There is a dearth of carefully controlled experimental data bearing upon the problem. Many observations upon infrahuman subjects suggest a very high degree of retention over long periods of time. Thus Davis (19) says that some of his raccoons showed but slight loss of ability to operate fastenings on puzzle boxes after more than a year without practice. With human subjects who have been trained in typewriting, Book (4) finds a high degree of retention after periods of six months and of one year during which the subject had no practice. Table 16 gives the results for one subject. As Book presents the results in the table, the first six months' interval shows that the average number of correct strokes made in the 10-day test is less than for the last 10-day period of the regular training test. The percentage of error was also greater. One year later the average for a period of equal length is even better than in the corresponding period 1.5 years before, although the error score remains high. At the beginning of each retest period the score is below what it was at the close of the last test. However, the relearning is accomplished within three or four 10-minute intervals. Book accounts for the high records made on the retests as follows: "The increase in score shown by our second memory series was due, so far as we could make out, rather *to the disappearance, with the lapse of time, of numerous psycho-physical difficulties, interfering associations, bad habits of attention, incidentally acquired in the course of learning, interfering habits and tendencies, which, as they faded, left the more firmly established typewriting associations free to act*" (4, p. 107). Results which in general are similar to those of Book have been found by Swift (94) and Bourdon (5) among others. Book's explanation may indicate one of the factors responsible for the high retention and rapid relearning. In addition one must indicate that the following factors may have played an important rôle: (1) The two subjects may have applied themselves in the retention tests with a vigor not present in the last of the training series. (2) The materials typed in the retest series may not have been so difficult as that of the original series. (3) There was a very great amount of overlearning of many of the responses involved. This overlearning would account for a high degree of retention and would make the results non-comparable with the work on nonsense syllables where there is a minimum of overlearning. And (4) Book's subjects had not reached their limits of performance during the training series and indeed were steadily improving. Under these conditions relearning may occur very quickly and the total score on the retest period may be superior to that of the final section of the training series. This problem needs much additional investigation before it can be adequately understood.

McGeoch and Melton (57) have made a preliminary report of a direct experimental comparison of the retention of a non-verbal response and of a verbal response with human subjects. They compared the retention of three maze habits after seven days with the retention of 8-, 12-, and 16-syllable lists after the same period of time. Mazes and lists were learned to the point of one correct repetition. Retention was tested by the savings method. "In terms of saving in trials, the nonsense syllables are uniformly better retained than are mazes. This superiority ranges from 30 to 60 per cent. In terms of saving in time, neither problem is uniformly superior. Five out of nine comparisons favor the mazes. In terms of saving in errors, the mazes excel in seven out of nine comparisons. In no case, save in trials where the syllables are very superior, are the differences great" (57, p. 144). This experiment which contributes dependable data on the comparative retentions of verbal and non-verbal habits offers no support for the view above mentioned that non-verbal habits persist longer than verbal ones. Probably the chief reason for the slow loss of non-verbal habits which is so often noted is the great amount of overlearning which they receive. The experiment by McGeoch and Melton suggests very strongly that any two responses learned to the same degree of proficiency will be equally well retained if the interpolated activities and the conditions of retesting do not favor one response to the detriment of the other and if the two responses are equally novel to the subjects tested. This generalized conclusion should receive special and detailed investigation.

Reminiscence

The loss of retention, as it has been traced by Ebbinghaus and his successors, shows a progress which is initially rapid and which then becomes less and less as time elapses. In 1913 Ballard (1) published the results of an experiment upon partially learned poetry with children as the subjects. The curves of retention secured from this work were radically different from anything heretofore described. The experiment was conducted as follows: A typewritten copy of the poetry to be learned was given each child. The experimenter then read the selection, and for 15 minutes the children were allowed to study the typewritten copies. These were then collected, and the children were told to write as much of the poetry as they could remember. Nothing was said about a second test; but at some predetermined later time, another written reproduction was called for. Figures 17 and 18 show typical results. The amount of immediate recall is taken as 100% and subsequent recalls are rated with this as a base. Each point on the curves represents a separate group of subjects. One group, i.e., was tested after one day; another group, after two days; etc. These curves do not mean that all children reproduced more after a lapse of time than they did immediately. Table 17 shows the actual percentages of children who reproduced more, less, or the same amount as they did immediately after study. The maximum reproduction for the six-year-olds is after an interval of three days; for 12-year-olds, after two days; and for 21-year-olds, immediately after study. Reminiscent items, i.e., items not contained in the immediate reproduction, may be present even though the total score on the second test is not so high as that on the first. The second reproduction is, therefore,

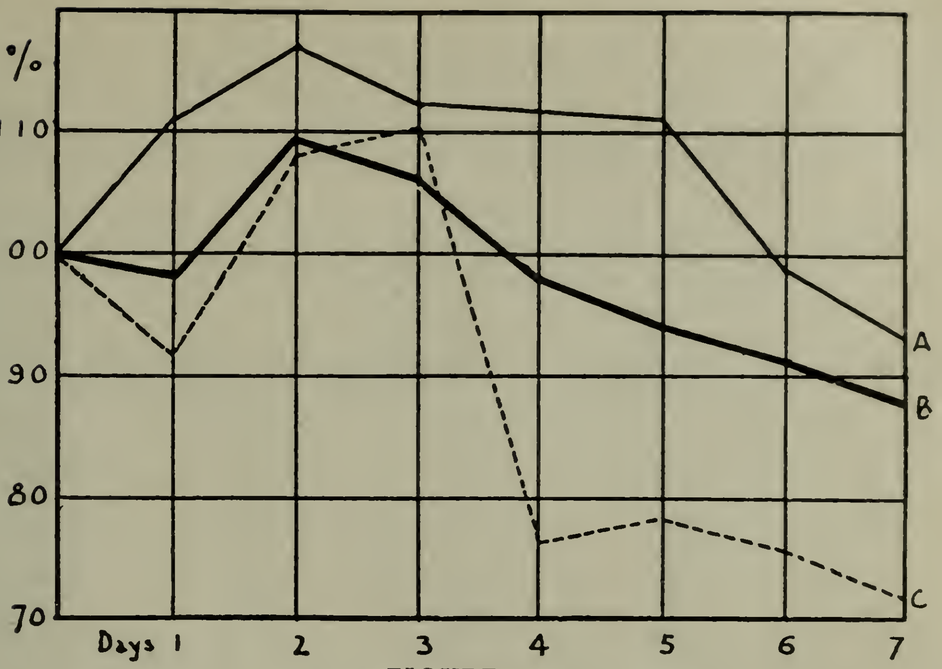


FIGURE 17

Memory Curves for Children of About 12 Years of Age

A is for the recall of "The Wreck of the Hesperus," *B*, of "The Ancient Mariner"; and *C*, of nonsense verses. (From Ballard, *Brit. J. Psychol., Monog. Suppl.*, 1913, 1.)

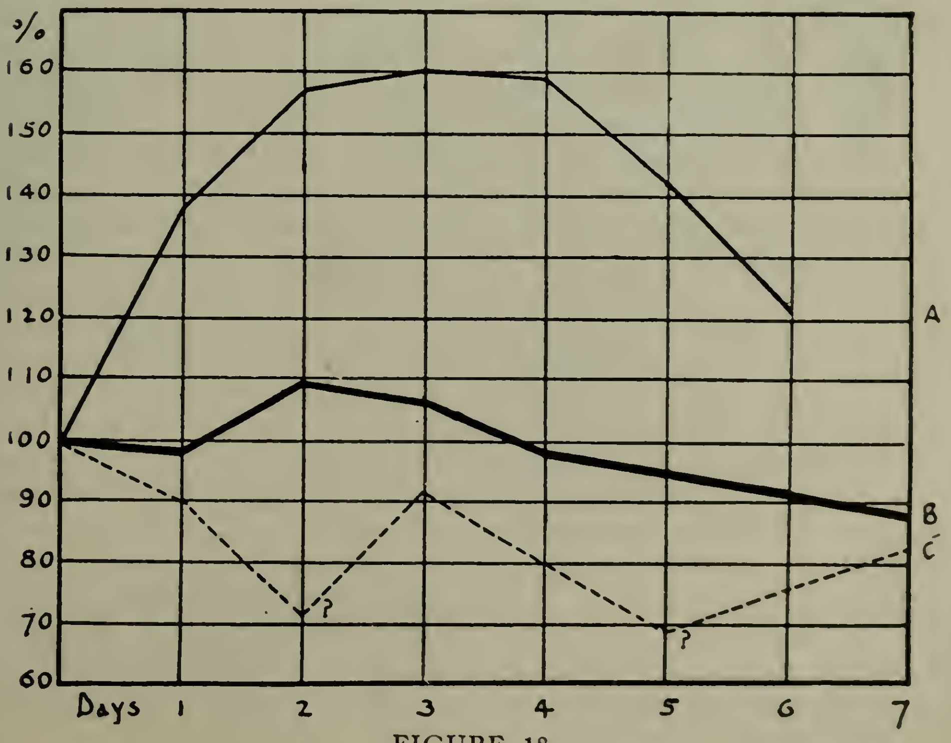


FIGURE 18

Memory Curves for Children of (*A*) about 6 Years Old, (*B*) about 12, and (*C*) about 21

(From Ballard, *Brit. J. Psychol., Monog. Suppl.*, 1913, 1.)

not solely determined by the first, a conclusion supported by Nicolai's (62) findings when the effects of the practice secured on the first reproduction were controlled. The influence of fatigue upon the initial test has also been ruled out as an explanatory factor. Luh (55) has failed to secure the phenomenon with adult subjects tested with nonsense syllables where learning was incomplete and where different methods of measuring retention were employed. Williams (116) and Huguenin (32) confirmed Ballard's work in extended studies with children where the material again was incompletely learned. In spite of the fact that Brown (7) found some signs of reminiscence in adults, the weight of evidence so far would indicate that age is the important accompanying condition of the phenomenon. Ballard found that the younger children and the defective children learned least and showed the greatest gain on the second reproduction.

TABLE 17

THE PERCENTAGE OF CHILDREN SHOWING REMINISCENCE AFTER VARIOUS INTERVALS OF TIME

(From Ballard)

Interval in days	Number of children	Percentage of children who after the interval remembered:		
		more	same	less
1	644	45.5	14.9	39.6
2	841	55.3	13.9	30.8
3	646	47.5	13.3	39.2
4	1086	40.6	15.1	44.3
5	657	42.2	13.1	44.7
6	660	30.	14.1	55.9
7	658	28.4	13.4	58.2

No adequate explanation for reminiscence is at present established. Ballard assumed that the neural processes involved in learning continue to set for an appreciable interval of time after the learning period has ended. "Reminiscence may be said to be due to the inertia of the nervous system, which does not yield to an influence at once. Nor does the inertia stop yielding at once. Evidence of this positive change is afforded by the phenomena of retro-active amnesia, and by the increase of skill in bodily activities that takes place during long periods of intermission of practice. While it is admitted that many of the facts seem to fit into either the theory of inhibition or the theory of neural growth, it seems to the writer that the latter is the only theory that adequately explains the whole of the phenomena with which we have been dealing" (1, p. 82). Huguenin and Ballard both, however, indicate that some items are lost as well as some recovered in the second test. Ballard's theory would account for the recovery but not for the loss. Brown regards the appearance of individual items in the second recall as partly a matter of chance and partly as a result of the fixating effect of the first recall. An adequate explanatory theory must await future experimentation.

Transfer of Training

No stimulus-response coordination can be acquired uninfluenced by the action system already possessed by the subject. Indeed it seems probable to many investigators that all new coordinations are merely the regroupings of simple responses which the organism has inherited. We have already considered the effect of training upon the retention of *previously* established habits, and we are now to discuss the effect of training upon the *subsequent* formation of habits. This effect may be positive or negative, that is, the formation of a stimulus-response coordination may be either facilitated or hindered by the training which has preceded it. In the former case we speak of habit transfer, and in the latter, of habit interference. The standard method of investigating such phenomena utilizes two groups of subjects of equal ability. One group is trained for a certain period of time on response no. 1, and then both groups are trained on response no. 2. A comparison of the records made by the two groups on response no. 2 will reveal whether or not the preliminary training has appreciably influenced the subsequent learning. Such comparisons are usually made in terms of the total time, trials, errors, or achievement of the groups, although they may also be made in terms of the learning curves. We shall first present typical experiments where non-verbal habits are concerned and then outline one or more experiments dealing with verbal habits.

Let us begin with observations made on cross-education. Investigations by Scripture (84), Davis (20), Thorndike and Woodworth (104), Starch (89), Ewert (23), and Bray (6) have shown that the training of responses made on one side of the body results in an increased facility of response on the opposite side of the body. Scripture studied strength of grip; Davis, quickness of tapping with the hands and feet; Thorndike and Woodworth, accuracy of tapping with the hands; Swift, ball-tossing; and Starch, Ewert, and Bray, mirror tests. Ewert used two groups of subjects having approximately equal ability as rated in terms of the Otis behavior test. These two groups were tested in tracing a star in the mirror apparatus under the following conditions:

Training group (26 subjects, average Otis score 164)
  (*a*) One trial with the non-preferred hand
  (*b*) Fifty trials with the preferred hand
  (*c*) One trial with the non-preferred hand
Control group (26 subjects, average Otis score 154)
  (*a*) One trial with the non-preferred hand
  (*b*) Interpolated rest, one hour
  (*c*) One trial with the non-preferred hand

Table 18 shows that the training group made 82% and 76% improvement in time and errors respectively in tracing with the non-preferred hand after an hour's practice with the preferred hand, while the control group made only 46% and 55% improvement after an equivalent rest period. The approximate amount of transfer from the training of the preferred hand is thus 36% for time and 21% for errors.

TABLE 18
RESULTS OF CROSS-EDUCATION IN MIRROR DRAWING
(After Ewert)

	Training group		Control group	
(*a*) 1st trial	237.7 sec.	86.6 errors	256.6 sec.	101 errors
(*c*) 2nd trial	43.6 sec.	20.7 errors	135.0 sec.	44.8 errors
% of improvement	82	76	46	55
gain due to training	36%	21%		

Hunter and Yarbrough (36), Pearce (66), and Hunter (34) have shown that habits established in the rat may interfere greatly with the acquisition of subsequent habits. In the last-mentioned study, rats were trained in a T-shaped discrimination box to go to the right side when a light was presented and to the left side for darkness. When this habit (habit no. 1) was mastered, the rats were trained to go to the right for

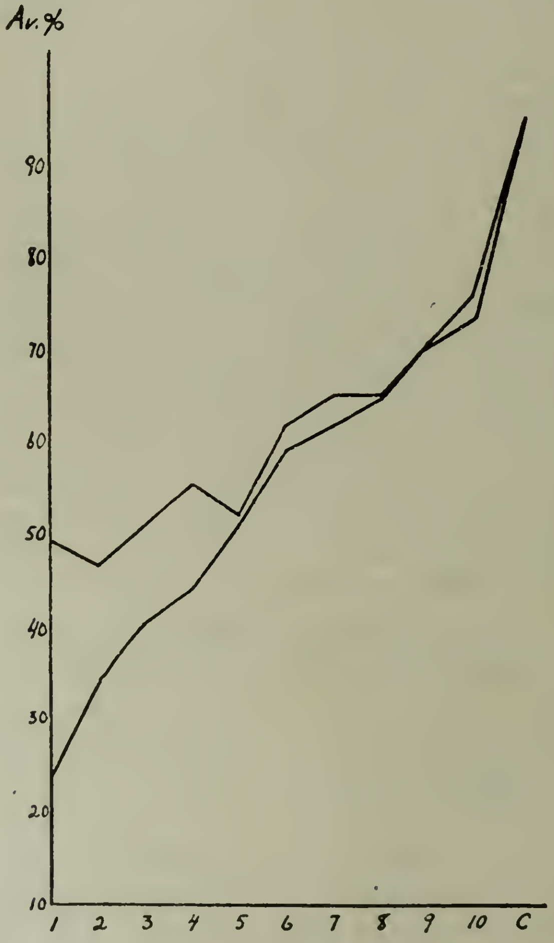


FIGURE 19
VINCENT CURVES TO SHOW THE LOCUS OF INTERFERENCE BETWEEN HABIT NO. 1 (UPPER CURVE) AND HABIT NO. 2 (LOWER CURVE)
(From Hunter, *J. Comp. Psychol.*, 1922, 2)

darkness and to the left for the light. An average of 286 trials was required to master habit no. 1 and an average of 603 trials to master habit no. 2 which, when not preceded by habit no. 1, was no more difficult than that habit. Figure 19, which gives Vincent learning curves for the two habits, shows that the locus of the interference was in the first one-half of the learning process.

Wylie's work (121) on white rats utilized avoiding reactions to different visual, auditory, and electrical stimuli. When the rat had learned to avoid one stimulus, another was substituted and the training was continued. In each case subsequent learning profited by the previous training. Hunter (33), in repeating this work, found that rats trained to make a given response to an auditory stimulus would, without additional training, make the same response to a visual stimulus. Positive transfer was thus indicated where all of the stimulating conditions were identical save one, the sound or the light.

Webb (112) and Wiltbank (117) have made important studies of the influence of learning one maze upon the mastery of later maze habits. In practically all cases transfer has been demonstrated. Even when the second maze requires a reversal of the turns made in the first, Hunter (34) has found transfer when the two learning records are compared in terms of total time, errors, and trials.

The early work on cross-education and most of the studies on the transfer of training using verbal material have been interpreted with reference to the educational theory of formal discipline. Formal discipline assumes that the explanation of positive transfer lies in the general improvement of behavior capacities and, conversely, that training in such forms of behavior as mathematics, logic, and Latin will improve the subject's general capacity to analyze and remember. The opponents of this doctrine explain transfer in terms of the training of elements which are identical in the forms of response concerned, and they base this interpretation upon experiments which show that training a given response results not in a general spread of improvement but in gains in a limited number of responses. The extensive literature on the subject is summarized by Thorndike (98) and Coover (17).

Thorndike and Woodworth (104) were the first to present experiments showing a very limited spread of improvement as a result of training in discriminative responses. The authors conclude as follows: "Improvement in any single mental function need not improve the ability in functions commonly called by the same name. It may injure it." "Improvement in any single mental function rarely brings about equal improvement in any other function, no matter how similar, for the working of every mental function-group is conditioned by the nature of the data in each particular case." "The general consideration of the cases of retention or of loss of practice effect seems to make it likely that spread of practice occurs only where identical elements are concerned in the influencing and influenced function" (104, p. 250). Somewhat greater spread of improvement was found by Fracker (24), who worked with visual and auditory stimuli, while the important study by Sleight (88) upon a variety of verbal habits shows almost no spread of improvement outside of the responses practiced.

Judd (40) has given a very brief description of an experiment which indicates that the verbal instruction which is given to the subject can aid in the mastery of a non-verbal habit, a result supported by one of Bray's (6) experiments.

"Two groups of pupils in the fifth and sixth grades were required to hit with a small dart a target which was placed under water. The difficulty of hitting the target arises, of course, from the deflection which the light suffers thru diffraction. . . . In this experiment one group of boys was given a full theoretical explanation of refraction. The other group of boys was left to work out experience without theoretical training. These two groups began practise with the target under twelve inches of water. It is a very striking fact that in the first series of trials the boys who knew the theory of refraction and those who did not, gave about the same results. That is, theory seemed to be of no value in the first tests. All the boys had to learn how to use the dart, and theory proved to be no substitute for practise. At this point the conditions were changed. The twelve inches of water were reduced to four. The difference between the two groups of boys now came out very strikingly. The boys without theory were very much confused. The practice gained with twelve inches of water did not help them with four inches. Their errors were large and persistent. On the other hand, the boys who had the theory fitted themselves to four inches very rapidly. Their theory evidently helped them to see the reason why they must not apply the twelve-inch habit to four inches of water" (40, pp. 36-37).

Poffenberger (73) carried out a series of tests upon control and training groups using the following tests: color naming, form naming, opposites, adjective-noun test, cancellation, and arithmetic. Table 19 summarizes the results. The last column indicates the presence or absence and the character of the transfer found when the figures in columns two and three are considered in the light of their probable errors.

We are still far from a satisfactory explanation of the facts of transfer. At present the theory of identical elements seems most adequate, and yet it is all but impossible to bring the theory to a rigorous experimental test. Perhaps work on conditioned reflexes may finally give us the explanatory key to the type of behavior interrelationships described in the present section.

The Relation of Initial and Final Abilities

This problem of the relation of initial ability to final ability has two important subdivisions: (1) What effect does training have upon the differences in ability revealed by individual subjects during the initial periods of the learning process? And (2) do the subjects used in a learning experiment tend to retain, at the close of training, the same relative ranks which they had at the beginning of training?

TABLE 19

Experiments upon Transfer with Verbal Responses
(After Poffenberger)

Type of experiment	Control group	Training group	Type of transfer
	Gross gain from 1st to last test	Gross gain from 1st to last test	
Influence of:			
Color naming upon form naming	6.0±1.2	—2.2±3.2	0
Opposites test upon adjective-noun test	15.6±3.2	—2.0±2.2	negative
Cancellation of 3 & 5 upon cancellation of groups containing 3 & 5	11.5±4	23.7±3.2	positive
Cancellation of 3 & 5 upon cancellation of groups containing 4 & 7	15.9±1.8	15.0±1.6	0
Addition upon substraction	15.1±4.8	8.8±0.7	0
Addition upon multiplication	29.7±3.8	11.4±5.3	negative
Addition upon division	19.5±4.3	21.3±6.3	0

Let us comment briefly upon this latter question first. The method of investigation is clear and relatively uncomplicated. Correlations are made between the initial and final performances. In order to promote the reliability of the initial and final records chosen, one may average the first two or three trials for the record of the initial standing of a given subject and average the records of the last two or three trials for the record of final performance. If the correlation secured is significant, we can predict final performance on the basis of initial scores. Where the correlation is positive, individuals who rank high initially will also tend to rank high finally. With negative correlations those subjects who rank high initially will tend to rank low finally. Many investigators have made correlations of the type suggested where the subjects have been trained on a great diversity of responses such as addition, multiplication, cancellation of letters on a page, color naming, typewriting, and running the maze. The correlations in the great majority of cases are positive, although the P.E.'s, and consequently the dependability of the coefficients, vary greatly. [Kincaid (42) has summarized results of this type from twenty-four different studies.] We may generalize the results as follows: the relationship between initial and final performances will in all probability be positive; but the closeness of the relationship, and therefore the degree to which valid predictions can be made, will vary from high to low and must be determined for each particular experimental setting.

Let us now turn to a consideration of the first question raised in the present section. What is the result of training upon individual differences? Do these differences increase, decrease, or remain the same? The sociological and educational significance of this problem is clear and important.

Do social pressure and school training tend to make individuals more nearly alike or do they serve to accentuate the differences between individuals? In a superficial sense and from a qualitative point of view, social conditions may result in a decrease of individual differences: immigrants become Americanized; children all learn the responses of reading, writing, and arithmetic; a code of moral behavior is known, even if not practiced, by the members of a group. Our interest in the problem, however, lies in another direction. Let us grant that 10 individuals have become more alike than they were by virtue of the fact that they are now all typists or calculators whereas formerly only some of them had these accomplishments. The quantitative and experimental problem which concerns us is this: Will the differences in skill which separate the 10 subjects one from the other in their initial efforts at typing or calculating increase, decrease, or remain the same after all subjects have received the same amount of training in typing or calculating?

Many different methods have been used in the attempted solution of this problem. These methods will be found listed with other items in column 1 of Table 20, and the results of the methods plus some of the data upon which they are based will be found in the later columns of the table. The results of this table seem to indicate, so far as the data here concerned are valid for an examination of the question, that individual differences in performance tend to decrease with practice. Where the standard deviation, S.D., is larger finally than it is initially, the amount of variability about the mean has increased. Since, however, the initial and final means are different, the coefficient of variability, S.D./Av., gives a more comparable measure of the spread of performance. "Ratio worst/best" is secured by dividing the worst initial performance by the best initial performance. This is then compared with a similar ratio from the final performances. "Gross gain" is secured by subtracting the initial score from the final score. Where the gross gain by the poorest 25% of the subjects was less than that made by the best 25% of the subjects, the letter *L* is used in the table. The same designation is used for the results secured by comparing "per cent gain."

Although Kincaid's tabulation and analysis point definitely to the conclusion that the differences in individual performance decrease with practice, we are not justified in accepting such a conclusion as final. As Chapman (15) and Stoddard (92) in particular have noted, certain factors which are vital and indispensable for a solution of the problem have been neglected by experimenters. (1) The subjects used have not only usually been too few in number but they have not been selected with reference to learning ability nor can we be reasonably sure that adequate incentives have been present, particularly in the final stages of training where improvement is most difficult. Behavior tests of the Binet or U. S. Army type give a rough measure of ability to profit by training. The suggestion is obvious, therefore, that experiments should be planned in such a way that different learning abilities should not complicate the results. (2) Account must be taken of the arbitrary character of the

TABLE 20

Data on Individual Differences as Affected by Training
(From Kincaid)

	Thorndike '08 Multiplication	Chapman '14 Addition	Chapman '14 Multiplication	Chapman '14 Color naming	Chapman '14 Opposites	Chapman '14 Cancelling 2's	Chapman '14 Cancelling 3's	Thorndike '16 Cancelling 2's	Thorndike '16 Cancelling 3's	Thorndike '16 Addition	Thorndike '16 Multiplication	Thorndike '10 Addition
Average (beginning)	261	65	43	201	68	91	95	113	128	61	61	159
Average (end)	521	84	75	222	117	110	111	173	162	108	189	231
S.D. beginning (B)	96	19	14	25+	14	12	11	15	15	16	17	41
S.D. end (E)	217	22	21	25—	17	15	16	18	13	23	76	66
Greater S.D.	E	E	E	B	E	E	E	E	B	E	E	E
S.D./Av. (beginning)	.37	.29	.33	.13	.21	.13	.12	.13	.12	.26	.28	.26
S.D./Av. (end)	.42	.26	.28	.11	.15	.14	.14	.10	.08	.21	.40	.29
Greater S.D./Av.	E	B	B	B	B	E	E	B	B	B	E	E
Ratio wosrt/best (beginning)	.22	.38	.20	.60	.42	.60	.67	.66	.65	.40	.35	.41
Ratio worst/best (end)	.21	.39	.25	.68	.58	.56	.55	.76	.78	.45	.25	.30
Greater ratio	B	E	E	E	E	B	B	E	E	E	B	B
Av. highest 25% beginning	397	93	63	189	87	105	109	133	145	82	81	213
Av. highest 25% end	703	116	97	245	136	110	129	175	171	126	231	287
Av. lowest 25% beginning	158	42	26	131	48	85	78	95	109	42	38	109
Av. lowest 25% end	316	63	52	189	105	92	70	166	152	84	133	162
Gross gain highest 25%	306	23	34	56	49	5	20	42	26	44	150	74
Gross gain lowest 25%	158	21	26	58	57	7	12	71	43	42	95	53
Greater gross gain	L	L	L	L	L	L	H	L	L	H	H	H
% gain highest 25%	.77	.25	.54	.30	.56	.05	.18	.32	.18	.54	1.85	.35
% gain lowest 25%	1.00	.50	1.00	.44	1.19	.08	.15	.75	.39	1.00	2.50	.49
Greater % gain	L	L	L	L	L	L	H	L	L	L	L	L
Correlation between initial performance and gross gain	.17	.34	.41	−.14	−.19	.05	.30	−.62	−.09	.27	.20	.05
P.E. of correlation	.12	.14	.13	.15	.14	.15	.14	.10	.15	.14	.14	.15
Correlation between initial per. and % gain	−.26.	−.43	−.38	−.36	−.55	−.24	.15	−.89	−.58	−.56	−.18	−.22

TABLE 20 (*continued*)

	Whitely '11 Maze	Whitely '11 Multiplication	Whitely '11 Cancelling A's	Thorndike '16 Typewriting	Woodrow '17 Form sorting, F.M.	Woodrow '17 Form sorting, Normal	McCall '16 Cancelling S's	McCall '16 Cancelling 3's	McCall '16 Addition	McCall '16 Copying addresses	Kincaid Braille writing	Kincaid Dart throwing
Average (beginning)	580	363	205	205	121	122	43	90	31	11	102	193
Average (end)	781	698	379	336	175	176	72	137	57	26	543	250
S.D. beginning (B)	172	198	56	84	17	16	22	27	14	4	29	73
S.D. end (E)	193	258	115	60	32	22	15	24	29	6	92	77
Greater S.D.	E	E	E	B	E	E	B	B	E	E	E	E
S.D./Av. (beginning)	.30	.55	.27	.41	.14	.13+	.51	.30	.45	.36	.29	.38
S.D./Av. (end)	.25	.37	.30	.18	.18	.13−	.21	.18	.51	.23	.17	.31
Greater S.D./Av.	B	B	E	B	E	B	B	B	E	E	E	E
Ratio worst/best (beginning)	.49	.18	.40	.23	.54	.64	.35	.24	.14	.05	.32	.16
Ratio worst/best (end)	.71	.38	.42	.58	.48	.62	.26	.14	.03	.33	.50	.32
Greater ratio	E	E	E	E	B	B	B	B	B	E	E	E
Av. highest 25% beginning	734	594	278	327	142	149	59	123	51	16	140	283
Av. highest 25% end	929	1043	558	427	195	192	84	148	94	31	607	301
Av. lowest 25% beginning	458	205	144	129	97	102	17	61	16	6	70	96
Av. lowest 25% end	695	579	294	278	156	153	59	130	31	22	505	188
Gross gain highest 25%	195	449	280	100	53	43	25	25	43	15	467	18
Gross gain lowest 25%	237	374	150	149	59	51	42	69	15	16	435	92
Greater gross gain	L	H	H	L	L	L	L	L	H	L	H	L
% gain highest 25%	.27	.76	1.01	.31	.37	.29	.42	.20	.84	.94	3.34	.06
% gain lowest 25%	.52	1.82	1.04	1.16	.61	.50	2.47	1.13	.94	2.67	6.21	.96
Greater % gain	L	L	L	L	L	L	L	L	L	L	L	L
Correlation between initial performance and gross gain	−.23	.18	.48	−.42	−.13	−.11	−.69	−.57	.57	−.10	.10	−.46
P.E. of correlation	.14	.14	.11	.13	.15	.15	.04	.05	.05	.05	.12	.09
Correlation between initial per. and % gain	−.42	−.33	−.27	−.79	−.36	−.40	−.55	−.81	.27	−.63	−.89	−.60

scale used for scoring. If one subject, A, receives an initial score of 10 and another subject, B, receives 20, B is not twice as good as A because zero on the arbitrary scale does not represent a zero of ability in the performance in question. The absolute or true zero may be 50 points below the arbitrary zero, and hence the true scores for A and B may be 60 and 70, respectively, rather than 10 and 20. In the absence of a determination of the true zero, no method of determining individual differences in terms of score ratios can properly be employed. Thurstone (105) has devised a method of determining the absolute zero which may be correct and applicable to the present problem. The method rests upon an absolute scaling of performance scores and upon the linear relationship which is claimed to exist between the standard deviation and the mean performance for various age groups. The use of gross gains from initial to final scores does not encounter this difficulty with the absolute zero, because this zero is common to both initial and final scores and is cancelled when one is subtracted from the other. (3) We come, therefore, to a consideration of the third important factor which has been neglected in the studies of individual differences. This factor is that of the inequality of units of achievement. The supposition has been that scores in addition, multiplication, or cancellation are composites of equal units. As Chapman (15, p. 230) wrote in 1925 concerning the use of gross scores, "The only assumption that this method makes is that the units at different points of the scale are equal." And yet this assumption can readily be shown to be untenable. Ordinarily the units of achievement in the early part of training are easier than those in the middle portion of the learning process, and the final units are the most difficult. The result, when graphically expressed, is a curve whcih changes rapidly at first and which then gradually levels off on the final plateau of achievement. As the limits of achievement are reached, and particularly as the performance approximates a physiological limit, the additional increments of accomplishment become more and more difficult to attain, and fewer and fewer subjects can master them. It therefore follows that the differences in individual performance may actually increase although the gross scores may indicate a decrease. Subject A may be 10 units better than subject B initially and only one point superior finally. And yet this one unit may mark an enormous difference between the two. At the close of 10 trials on a maze one rat may be able to run without error while another rat may make but one error. In certain mazes this difference in performance is not small but great because the one error that remains is the most difficult one to eliminate, indeed many subjects fail to eliminate it at all. We may choose still other illustrations of the same thing. To reduce one's golf score from 100 to 98 is not nearly so difficult as to reduce one's score 2 points below par. Five-tenths of a second clipped off the world's record for the 100-yard dash is a much more considerable achievement than to reduce the time from 20 secs. to 19.5 secs. for the same distance. In these cases in which improvement involves an approximation to the physiological limits of the organism, the work by Hill (29) on the physiology of achievement in athletics suggests the importance of finding some way to relate the difficulty of units to the capacities of the working organisms.

Before we can solve this question of the influence of training upon individual differences a practical method must be found for weighting the units of score in terms of their difficulty. If in addition we can secure probable determinations of the true zero, our work will be facilitated. The rôle of the prophet in science is not a secure one, and yet we hesitate but little in predicting that future investigations made in the light of the above requirements will show that individual differences in degree of skill increase rather than diminish with training.

The Laws of Learning

In closing our discussion of the learning process, brief reference must be made to one of the most important and yet one of the most baffling problems in the field. What is the mechanism of the learning process? What factors are operative in determining the acquisition and elimination of responses during the course of training? Many responses occur in the initial stages of habit formation of which some are incorporated into the final habit and others are not. Conventional theory states that the essential factors which determine this trend in behavior are the frequency and recency of the response plus the effect which follows upon the response. The theory pays little attention to current neurological investigations or to the work on conditioned reflexes because problems in these two fields have not been developed with reference to the laws of learning. Watson in 1914 placed exclusive emphasis upon frequency and recency. To these Carr (13) would add intensity, a variation of the factor of effect. Thorndike (97) in 1911 limited himself to frequency and effect and formulated their laws as follows: "The Law of Effect is that: *Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.*

"The Law of Exercise is that: *Any response to a situation will, other things being equal, be more strongly connected with the situation in proportion to the number of times it has been connected with that situation and to the average vigor and duration of the connections.*"

The law of recency, which is not included in Thorndike's formulation, assumes that the response which has been active most recently is more likely to be reinstated than one which is less recent. Since any one trial in learning will terminate with a correct response, this response will probably be called out earlier the next trial because of the lower threshold of its controlling neural mechanism.

The conventional laws of learning which we have just outlined are derived historically from the laws of association. They give a controlling rôle in the determination of behavior to factors which after all are merely descriptive of certain aspects of behavior, leaving essentially untouched the underlying and fundamental neural conditions.

The experimental investigation of the laws of learning has not as yet been pushed far enough to give us a satisfactory account of such phenomena as the elimination of blind alleys in a maze or the elimination of errors in speaking or typewriting. We can, however, refer to several experimental attempts that have been made to advance the problem. Efforts have been made to study the influence of frequency, recency, and effect upon learning in the maze. In the latter case the escape from the maze with the consequent stimuli from the food and eating constitute the effect. If this effect is influential in eliminating entrances into the blind alleys, then the blinds nearest the food box should be eliminated first and those at the entrance should be eliminated last. In general this is what does happen; but Peterson (71), in maze tests where the direction of response was determined by tossing a coin, has shown that this is in accordance with the laws of chance and that (72) frequency and recency would not tend to eliminate but to fix erroneous responses. Peterson's own theory is that the selection of responses which occurs in learning is the result of factors growing out of the completeness of response. This is essentially a Gestalt theory, the physiological aspects of which Peterson develops in terms of the drainage of neural activities away from those responses which do not fit into the consummatory processes, e.g., food-getting, which terminate the trial.

Thorndike (100) tested the influence of effect as follows: The subject was shown strips of paper, ranging in length from 3-27 cm., each of which was to be estimated in length. A standard strip 10 cm. long was always in the field of vision. Three series of estimates were given: (1) a series of 50 presentations in which the subject was not told whether he was right or wrong; (2) seven series in which the experimenter said right or wrong after each estimate; and (3) another series in which the experimenter said nothing. The effect factor was thus the "right" or "wrong," reward and punishment, from the experimenter. A control group of subjects were given the same experiment except that the seven series of training were mere repetition without the words right or wrong being spoken. The result was that the experimental group reduced its score on the third test by an average of 61%, while the control group raised their score in the third test 7% above what it was on the first test, i.e., they lost in performance rather than gained. Thorndike writes as follows: "These experiments are crucial as a demonstration that the consequences of a connection work back upon it to influence it. There is no difference between the 'aided' and the 'no aid' experiences save in the consequences of the connection. It is also evident that the consequences probably act on the connection directly, and not by leading the *S* to repeat it, or something like it, to himself" (100, pp. 215-6). This conclusion is in harmony with the evidence drawn from work on conditioned reflexes.

Thorndike in another experiment (101) has challenged the law of frequency upon which great emphasis has always been placed in explaining the fact of learning. The subjects were presented with a series of tasks such as estimating the lengths of lines and drawing lines of specified lengths or angles of specified sizes. A comparison was made between the records of the first part of each experiment and the records of the last part in order to determine whether the response made most frequently initially was also made still more frequently at the close. In no case did the most frequently made response profit significantly from its greater frequency. Thorndike says: "From these and other experiments we conclude that mere frequency of use of [a stimulus-response] does not strengthen it at the expense of weaker connections. . . . Greater relative frequency is not a selective force. To explain the facts of modifiability some influence from the consequences of a connection is required" (101, p. 18). We cannot follow Thorndike in rejecting frequency upon the basis of these experiments. The phenomenon of incidental memory has long familiarized us with the impotency of mere frequency of stimulation and response in producing learning, but it is not clear (although it may be true) that incidental memory needs only the introduction of an effect factor in order to have its status changed. In order to modify behavior significantly a stimulus-response must apparently dominate the organism at the time when it is repeated, dominate in some such manner as the old associationists meant when they said that things must be *experienced together* in order to be associated. Our comments are not meant to belittle Thorndike's experiment but to urge caution in its interpretation, whether that be negative or positive. The necessary experimental results are not as yet all in. Experiments on conditional reflexes make us think that frequency is one of the important factors underlying conditioning. Another vital factor is the temporal relation between the presentation of the effective and the non-effective stimulus. If, for example, the presence of food in the mouth causes saliva to flow and if it is desired to produce that result with an auditory stimulus, the latter must either be presented simultaneously with the food or slightly before the food. Presentation subsequent to the food will not result in the conditioned response. Even when the temporal relationship is optimal and when the effect factor of the food is present, a certain frequency of stimulation is necessary for the appearance of the conditioned response. The present chapter has not afforded an opportunity for a detailed presentation of the work upon conditioned reflexes. Had such an opportunity been available, the conclusion would have been suggested that all learning is fundamentally of the conditioned-reflex type and that the analysis of the detailed mechanics of the learning process must be based upon detailed physiological experimentation.

REFERENCES

1. Ballard, P. B. Obliviscence and reminiscence. *Brit. J. Psychol., Monog. Suppl.*, 1913, **1**, No. 2.
2. Batson, W. H. Acquisition of skill. *Psychol. Monog.*, 1916, **31**, No. 3.

3. Bird, C. The effect of maturation upon the pecking instinct of chicks. *Ped. Sem.*, 1926, **33**, 212-234.
4. Book, W. F. The psychology of skill. (Reprinted from *Univ. Montana Stud.*, 1908, vol. 1.) New York, 1925. Pp. 257.
5. Bourdon, B. Recherches sur l'habitude. *Année psychol.*, 1901, **8**, 327-340.
6. Bray, C. W. Transfer of learning. *J. Exper. Psychol.*, 1928, **11**, 443-467.
7. Brown, W. To what extent is memory measured by a single recall? *J. Exper. Psychol.*, 1923, **6**, 377-382.
8. ——————. Effects of interval on recall. *J. Exper. Psychol.*, 1924, **7**, 469-474.
9. ——————. Whole and part methods of learning. *J. Educ. Psychol.*, 1924, **15**, 229-233.
10. Bryan, W. L., & Harter, N. Studies in the physiology and psychology of the telegraphic language. *Psychol. Rev.*, 1897, **4**, 27-53; 1899, **6**, 345-357.
11. Calkins, M. W. Association. *Psychol. Monog.*, 1896, **1**, No. 2.
12. Carmichael, L. A further study of the development of behavior in vertebrates experimentally removed from the influence of external stimulation. *Psychol. Rev.*, 1927, **34**, 34-47.
13. Carr, H. A. Principles of selection in animal learning. *Psychol. Rev.*, 1914, **21**, 157-165.
14. ——————. Distribution of effort. *Psychol. Bull.*, 1919, **16**, 26-28.
15. Chapman, J. C. Statistical considerations in interpreting the effects of training on individual differences. *Psychol. Rev.*, 1925, **32**, 224-234.
16. Clements, F. E. The effect of time on distance discrimination in the albino rat. *J. Comp. Psychol.*, 1928, **8**, 317-324.
17. Coover, J. E. Formal discipline from the standpoint of experimental psychology. *Psychol. Monog.*, 1916, **20**, No. 3.
18. Cummins, R. A. Improvement and the distribution of practice. *Teach. Coll. Contrib. Educ.*, 1919, No. 97.
19. Davis, H. B. The raccoon: a study in animal intelligence. *Amer. J. Psychol.*, 1907, **18**, 447-490.
20. Davis, W. W. Researches in cross-education. *Stud. Yale Psychol. Lab.*, 1898, **6**, 6-50; 1900, **8**, 64-108.
21. Ebbinghaus, H. Memory. (Trans. by Ruger and Bussenius.) New York: 1913. Pp. 123.
22. Ephrussi, P. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zsch. f. Psychol.*, 1904, **37**, 56-103, 161-224.
23. Ewert, P. H. Bilateral transfer in mirror-drawing. *Ped. Sem.*, 1926, **33**, 235-249.
24. Fracker, G. C. On the transference of training in memory. *Psychol. Monog.*, 1908, **9**, No. 2.
25. Gates, A. I. Recitation as a factor in memorizing. *Arch. Psychol.*, 1917, **6**, No. 40.
26. Gordon, K. Class results with spaced and unspaced memorizing. *J. Exper. Psychol.*, 1925, **8**, 337-343.
27. Harden, L. M. A quantitative study of the similarity factor in retroactive inhibition. *J. Gen. Psychol.*, 1929, **2**, 197-221.
28. Heine, R. Ueber Wiedererkennen und Rückwirkende Hemmung. *Zsch. f. Psychol.*, 1914, **68**, 161-236.
29. Hill, A. V. Muscular movement in man: the factors governing speed and recovery from fatigue. New York, 1927. Pp. 104.
30. Hobhouse, L. T. Mind in evolution. New York, 1901. Pp. 415.

31. Hubbert, H. B. The effect of age on habit formation in the albino rat. *Behav. Monog.*, 1915, **2**, No. 11.

32. Huguenin, C. Reviviscence paradoxale. *Arch. de. psychol.*, 1914, **14**, 379-383.

33. Hunter, W. S. Some notes on the auditory sensitivity of the white rat. *Psychobiol.*, 1918, **1**, 339-351.

34. ———. Habit interference in the white rat and in human subjects. *J. Comp. Psychol.*, 1922, **2**, 29-59.

35. ———. Human behavior. Chicago, 1928. Pp. 355.

36. Hunter, W. S., & Yarbrough, J. U. The interference of auditory habits in the white rat. *J. Animal Behav.*, 1917, **7**, 49-65.

37. Jenkins, J. G., & Dallenbach., K. M. Obliviscence during sleep and waking. *Amer. J. Psychol.*, 1924, **35**, 605-612.

38. Jennings, H. S. The behavior of lower organisms. New York, 1906. Pp. 366.

39. Jost, A. Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Zsch. f. Psychol.*, 1897, **14**, 436-472.

40. Judd, C. H. The relation of special training to general intelligence. *Educ. Rev.*, 1908, **36**, 28-42.

41. Katzaroff, D. Le rôle de la récitation comme facteur de la mémorisation. *Arch. de psychol.*, 1908, **7**, 224-259.

42. Kincaid, M. A study of individual differences in learning. *Psychol. Rev.*, 1925, **32**, 34-53.

43. Kjerstad, C. L. The form of the learning curves for memory. *Psychol. Monog.*, 1919, **26**, No. 5.

44. Knors, C. Experimentelle Untersuchungen über den Lernprozen. *Arch. f. d. ges. Psychol.*, 1910, **17**, 297-362.

45. Koffka, K. The growth of the mind. New York, 1924. Pp. 382.

46. Köhler, W. The mentality of apes. New York, 1925. Pp. 342.

47. ———. Gestalt psychology. New York, 1929. Pp. 403.

48. Krueger, W. C. F. The effect of overlearning on retention. *J. Exper. Psychol.*, 1929, **12**, 71-78.

49. Kühn, A. Ueber Einprägung durch Lesen und durch Rezitieren. *Zsch. f. Psychol.*, 1914, **68**, 396-481.

50. Lakenan, M. E. The whole and part methods of memorizing poetry and prose. *J. Educ. Psychol.*, 1913, **4**, 189-198.

51. Larguier des Bancels, J. Sur les méthodes de mémorisation. *Année psychol.*, 1901, **8**, 185-213.

52. Lashley, K. S. A causal factor in the relation of the distribution of practice to the rate of learning. *J. Animal Behav.*, 1917, **7**, 139-142.

53. ———. A simple maze: with data on the relation of the distribution of practice to the rate of learning. *Psychobiol.*, 1918, **1**, 353-367.

54. Liu, S. Y. The relation of age to the learning ability of the white rat. *J. Comp. Psychol.*, 1928, **8**, 75-86.

55. Luh, C. W. The conditions of retention. *Psychol. Monog.*, 1922, **31**, No. 3.

56. Lyon, D. O. The relation of quickness of learning to retentiveness. *Arch. Psychol.*, 1916, **5**, No. 34.

57. McGeoch, J. A., & Melton, A. W. The comparative retention values of maze habits and nonsense syllable. *Psychol. Bull.*, 1929, **26**, 144.

58. Meumann, E. The psychology of learning. (Trans. by Baird.) New York, 1913. Pp. 387.

59. Moseley, D. The accuracy of the pecking response in chicks. *J. Comp. Psychol.*, 1925, **5**, 75-97.

60. Müller, G. E., & Pilzecker, A. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zsch. f. Psychol.*, 1900, Erg., **1**. Pp. 300.

61. Neumann, G. Experimentelle Beiträge zur Lehre von der Ökonomie und Technik des Lernens. *Zsch. f. exper. Päd.*, 1907, **4**, 63-101.

62. Nicolai, F. Experimentelle Untersuchungen über das Haften von Gesichtseindrücken und dessen zeitlichen Verlauf. *Arch. f. d. ges. Psychol.*, 1922, **42**, 132-149.

63. Norsworthy, N. Acquisition as related to retention. *J. Educ. Psychol.*, 1912, **3**, 214-218.

64. Paterson, D. G. The Johns Hopkins circular maze studies. *Psychol. Bull.*, 1917, **14**, 294-297.

65. Pavlov, I. P. Conditioned reflexes. (Trans. by Anrep.) London, 1927. Pp. 430.

66. Pearce, B. D. A note on the interference of visual habits in the white rat. *J. Animal Behav.*, 1917, **7**, 169-177.

67. Pechstein, L. A. Whole *vs.* part methods in motor learning. *Psychol. Monog.*, 1917, **23**, No. 2.

68. Pentschew, C. Untersuchungen zur Ökonomie und Technik des Lernens. *Arch. f. d. ges. Psychol.*, 1903, **1**, 417-526.

69. Peterson, G. M. Negative acceleration with material of varying difficulty. *J. Exper. Psychol.*, 1928, **11**, 40-44.

70. Peterson, J. Experiments in ball-tossing: the significance of learning curves. *J. Exper. Psychol.*, 1917, **2**, 178-224.

71. ————. Frequency and recency factors in maze learning by white rats. *J. Animal Behav.*, 1917, **7**, 338-364.

72. ————. Learning when frequency and recency factors are negative. *J. Exper. Psychol.*, 1922, **5**, 270-300.

73. Poffenberger, A. T. The influence of improvement in one mental process upon other related processes. *J. Educ. Psychol.*, 1915, **6**, 459-474.

74. Pyle, W. H. Standards of mental efficiency. *J. Educ. Psychol.*, 1913, **4**, 61-70.

75. ————. The psychology of learning. Baltimore, 1921. Pp. 308.

76. Pyle, W. H., & Snyder, J. C. The most economical unit for committing to memory. *J. Educ. Psychol.*, 1911, **2**, 133-142.

77. Radossawljewitsch, P. Das Behalten und Vergessen bei Kindern und Erwachsenen. Leipzig, 1906.

78. Reed, H. B. Part and whole methods of learning. *J. Educ. Psychol.*, 1924, **15**, 107-115, 248-249.

79. Robinson, E. S. Some factors determining the degree of retroactive inhibition. *Psychol. Monog.*, 1920, **28**, No. 6.

80. ————. The similarity factor in retroaction. *Amer. J. Psychol.*, 1927, **39**, 297-312.

81. Robinson, E. S., & Darrow, C. W. Effect of length of lists upon memory for numbers. *Amer. J. Psychol.*, 1924, **35**, 235-243.

82. Robinson, E. S., & Heron, W. T. Result of variations in length of memorized material. *J. Exper. Psychol.*, 1922, **5**, 428-448.

83. Ruch, T. C. Factors influencing the relative economy of massed and distributed practice in learning. *Psychol. Rev.*, 1929, **35**, 19-45.

84. Scripture, E. W., Smith, T. L., & Brown, E. M. On the education of muscular control and power. *Stud. Yale Psychol. Lab.*, 1894, **2**, 114-119.

85. Shepard, J. F., & Breed, F. S. Maturation and use in the development of an instinct. *J. Animal Behav.*, 1913, **3**, 274-285.

86. Skaggs, E. B. The relative value of grouped and interspersed recitations. *J. Exper. Psychol.*, 1920, **3**, 424-446.

87. ————. Further studies in retroactive inhibition. *Psychol. Monog.,* 1925, **34**, No. 161.

88. Sleight, W. G. Memory and formal training. *Brit. J. Psychol.,* 1911, **4**, 386-457.

89. Starch, D. A. A demonstration of the trial and error method of learning. *Psychol. Bull.,* 1910, **7**, 20-23.

90. ————. Periods of work in learning. *J. Educ. Psychol.,* 1912, **3**, 209-213.

91. Steffens, L. Experimentelle Beiträge zur Lehre vom ökonomischen Lernen. *Zsch. f. Psychol.,* 1900, **22**, 321-382.

92. Stoddard, G. D. The problem of individual differences in learning. *Psychol. Rev.,* 1928, **35**, 479-485.

93. Stone, C. P. The age factor in animal learning: I. Rats in the problem box and the maze. *Genet. Psychol. Monog.,* 1929, **5**, 1-130.

94. Swift, E. J. Studies in the psychology and physiology of learning. *Amer. J. Psychol.,* 1903, **14**, 201-251.

95. ————. Beginning a language. (Garman Memorial Volume.) New York, 1906. Pp. 297-313.

96. Swift, E. J., & Schuyler, W. The learning process. *Psychol. Bull.,* 1907, **4**, 307-310.

97. Thorndike, E. L. Animal intelligence. New York, 1911. Pp. 297.

98. ————. The psychology of learning. New York, 1913. Pp. 452.

99. ————. Work, fatigue, and individual differences. New York, 1914. Pp. 408.

100. ————. The law of effect. *Amer. J. Psychol.,* 1927, **39**, 212-222.

101. ————. A fundamental theorem in modifiability. *Proc. Nat. Acad. Sci.,* 1927, **13**, 15-18.

102. Thorndike, E. L., *et al.* The measurement of intelligence. New York, 1926. Pp. 616.

103. ————. Adult learning. New York, 1928. Pp. 335.

104. Thorndike, E. L., & Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions. *Psychol. Rev.,* 1901, **8**, 247-261, 384-395, 553-564.

105. Thurstone, L. L. The absolute zero in intelligence measurement. *Psychol. Rev.,* 1928, **35**, 175-197.

106. Trow, W. C., & Sears, R. A learning plateau due to conflicting methods of practice. *J. Educ. Psychol.,* 1927, **18**, 43-47.

107. Ulrich, J. L. Distribution of effort in learning in the white rat. *Behav. Monog.,* 1915, **2**, No. 5.

108. Vincent, S. B. The function of the vibrissae in the behavior of the white rat. *Behav. Monog.,* 1912, **1**, No. 5.

109. Warden, C. J. Distribution of practice in animal learning. *Comp. Psychol. Monog.,* 1923, **1**, No. 3.

110. Watson, J. B. Kinaesthetic and organic sensations: their rôle in the reactions of the white rat in the maze. *Psychol. Monog.,* 1907, **8**, No. 2.

111. ————. Psychology from the standpoint of a behaviorist. Philadelphia, 1919. Pp. 429.

112. Webb, L. W. Transfer of training and retroaction. *Psychol. Monog.,* 1917, **24**, No. 3.

113. Whipple, G. M. Manual of mental and physical tests. 2 vols. Baltimore, 1914, 1915. Pp. 365, 336.

114. Whitely, P. The dependence of learning and recall upon prior mental and physical conditions. *J. Exper. Psychol.,* 1924, **7**, 420-428.

115. ————. The dependence of learning upon prior intellectual activities. *J. Exper. Psychol.,* 1927, **10**, 489-508.

116. Williams, O. A study of the phenomenon of reminiscence. *J. Exper. Psychol.,* 1926, **9**, 368-387.

117. Wiltbank, R. T. Transfer of training in white rats upon various series of mazes. *Behav. Monog.,* 1919, **4**, No. 1.

118. Winch, W. H. Should poems be learnt by school-children as 'wholes' or in 'parts'? *Brit. J. Psychol.,* 1924, **15**, 64-79.

119. Witasek, S. Ueber Lesen und Rezitieren in ihrer Beziehungen zum Gedächtnis. *Zsch. f. Psychol.,* 1907, **44**, 246-278.

120. Woodworth, R. S. A contribution to the question of "quick learning, quick forgetting." *Psychol. Bull.,* 1914, **11**, 58-59.

121. Wylie, H. H. An experimental study of transfer of response in the white rat. *Behav. Monog.,* 1919, **3**, No. 5.

122. Yerkes, R. M. The mental life of monkeys and apes. *Behav. Monog.,* 1916, **3**, No. 1.

123. ————. The mind of a gorilla. *Genet. Psychol. Monog.,* 1927, **2**, 1-193, 375-551.

CHAPTER 16

THE INDIVIDUAL IN INFANCY

ARNOLD GESELL
Yale University

INFANCY AND THE LIFE-CYCLE

Infancy is a flexible term. The biologist uses it to designate the whole span of immaturity. At the common law an infant becomes of age the first instant of the day preceding the 21st anniversary of his birth. From the standpoint of genetic psychology infancy is ordinarily restricted to the first stage of immaturity, but there is no agreement as to when this stage terminates.

Nor is there a stable, inclusive criterion for maturity. The suggestion that there is no appreciable growth of mental abilities after the age of 16 does not supply a true psychobiological limit to maturity. Various forms of mental maturation similar to those of babyhood may continue into the twenties and beyond. One of the tasks of developmental psychology is to establish the sequences and continuity of growth phenomena, whether of increase or of decline, throughout the whole life-cycle. Certain manifestations of abnormal mental deterioration, as well as of normal senectitude, must be brought into comparative, if not genetic relationship with the characteristics of infancy. The traits of maturity also are foreshadowed on an extensive scale by traits displayed in early life. Moreover, the dynamic organization of the adult may be lawfully influenced by the events and the patterns of the first formative growth stages. Infancy thus literally projects itself into subsequent years. For all these reasons, the psychology of infancy cannot stand apart in a peculiar province, but must be articulated with the psychologies of all sectors of the life-cycle and with the general psychology of mature man.

For chronological convenience one may regard the span of postnatal immaturity as embracing four periods of six years each. The first sexennium is the pre-school period, and terminates with the eruption of punctual sixth-year molars. The second sexennium is the period of preadolesence extending to the teens (and the second molars). The third and fourth sexennia, early and later adolescence, bring the youth to the borders of maturity, marked by the wisdom tooth.

In a restrictive sense the term infancy will be used in the present chapter to refer to the first two years of life. Special emphasis will be given to the first of these two years, but the whole pre-school period will be included in the discussion to conform with the planning of the volume. For further classification early infancy may be regarded as the perception-prehension period lasting about 40 weeks, followed by the language-locomotion period up to the age of two years. The first four weeks of infancy constitute the neonatal period.

These periods merge and are profoundly continuous in the organic unity of the life-cycle. The neonatal period is really but an extension of the prenatal period, and in cases of premature birth actually coincides in a measure with the latter.

One may think of infancy as an epoch in an individual life-cycle; but from the broad standpoint of psychobiology infancy constitutes an intricate trait-complex—a complex which has changed through the ages and is still in process of change. As such it is one of the most complicated end-products of evolution. Infancy itself was evolved to subserve the needs of growth. It is not simply an interval in which an individual manifests growth, but it is *a method of growth* by which the potentialities of the life-cycle are ultimately realized.

Infancy is therefore not purely a prologue or a metamorphosis. From the organismic viewpoint infancy plays a positive, persisting rôle in the attainment of the total life-career. The comparative psychology of infancy becomes most significant when it reveals the lawful correlation between a given kind of infancy and a given kind of life-cycle throughout a series of species. Such correlations are an index to the most general laws of growth. Even though these correlations have not yet been established, it is certain that they will yield to systematic comparative study. They are a necessary foundation for a general physiology of behavior development.

The Experimental Significance of Age

Time permeates every corner of physics, says Eddington in his lectures on *The Nature of the Physical World.* He is speaking of Astronomer Royal time, the metric of duration. The same kind of time surely permeates every corner of developmental phenomena. The very idea of growth calls up concepts of time-lapse and of time-consumption, and also of cycle and of age.

It takes time to grow. Development is profoundly conditioned by duration. Although we cannot suggest that development is caused by the clock or the calendar, it is scientifically important to recognize that development is a function of that special form of duration called age.

There has been a tendency to regard age simply as a formal device for dating or grouping the observations of infant development, but this does not do justice to the intrinsic significance of age. Age is more than a dating. Age is at once an orientational and a quantitative category. Any given age always represents a position in a cycle. For this reason the precise determination of true age becomes of considerable importance in the methodology of genetic study. This is particularly true of early infancy, when development is proceeding at such a rapid rate that small units of duration count heavily.

There is yet another consideration which adds to the significance of accurate age determination in the study of infant behavior, namely, the variability of the "times" at which children are born. Normally human gestation covers a period of 40 weeks, but birth with survival may exceptionally occur as early as 24 weeks and as late as 48 weeks, an enormous range of variation amounting to six lunar months. Variation with a range of three lunar months is common.[1] It has long been the practice to reckon a child's age from the time of birth, but this is a most naïve convention which mocks the dignity of the true zero.

Does this viewpoint place too much weight upon the concept of age? It may be urged that age is, after all, a matter of clock and calendar, and does not inhere in the substance of reality; that the speed of early growth, for example, is a function of temperature, of enzymes, of protein ration; that the truly fundamental factors in growth are chemical and physical and not in any sense temporal. Give the experimenter control of the fundamental factors, and he may stop growth in his laboratory animals, retard growth, accelerate it. He can convert a moment developmentally into a long-drawn cycle, and thus play tricks with the factor of age. He, and nature, too, has amply demonstrated the relativities of age.

But this line of reasoning does not undermine the scientific importance of the concept of age. The events of growth and of behavior are ordered; they fall into temporal sequences. There may be telescoping, elision, hastening, slowing; but the events will always appear, so to speak, in a time frame, and have a durational design which can be expressed in temporal terms. Theoretically, every genetic phenomenon can be dated. The quantitative analysis and the statistical study of such dates constitutes the study of age—leads to the elucidation of age in relation to individual differences. Inasmuch as genetic phenomena are time-linked, and are fundamentally conditioned by duration, accurate age data are fundamental for the expression of the patterns of growth, whether seen in cross-section, longitudinally, or biographically. Age norms, or age values, are necessary for the formulation of the trends of typical and of deviational characteristics.

Far from being a restrictive and formalistic concept, the critical use of age norms places a premium on the delineation of individual differences and favors the conjoint formulation of general trends and of specific departure from trends. Paradoxically, the systematic development of age norms leads to the possibility of the diagnosis of age itself, when age is not known. No adequate method has yet been devised for the determination of the post-conception age of the newborn child, a problem of special significance in the premature. It is conceivable that refined studies of preterm extra-uterine behavior, will lead to the definition of behavior indices which will furnish a sensitive means for calculating the age and maturity of the individual at birth.

---

[1] An instructive instance of this variation is to be found in no less a subject than Miss Millicent Shinn's niece who furnished the data for the often quoted *Notes on the Development of a Child.* The age value of many of the reported observations is apparently affected by postmaturity of birth, a complication to which Miss Shinn calls attention.

THE DEVELOPMENTAL ZERO

A discussion of the critical significance of age again leads back to an analysis of the relations of infancy to the human life-cycle. On its systematic side the developmental psychology of infancy is inevitably concerned with the age values of behavior and the behavior values of age. Even though there may be no genuine absolute unit of measurement, these values must be scaled from a common zero base.

The zero of chronological age presents no mathematical problem. It is assumed that chronological age is measured in units of duration, which are equal in an absolute or clock-time sense. The problem simply is: When shall we set the clock ticking? The law says at birth. Biology says before birth.

The "absolute" zero of development likewise presents no peculiar problem, if we posit that this zero must coincide with the clock zero, because developmental age and chronological age are by premise equated—and an ideal (normative) correspondence between chronological and developmental ages is postulated. The zero location may therefore advance or recede with the requirements and refinements of technique; or the mensurational aspects of the problem may be solved by establishing specific zeros for prehension, for speech, for altruism, for intelligence, etc.

In this sense, unless the problem is approached on purely mathematical grounds,[2] the problem of the ontogenetic zero is a relative one, subject to the varying needs of research and of quantitative statement. From our immediate point of view, the zero is significant, not because it is necessary for the establishment of a series of absolutely equal units, but because there must be a sufficiently fundamental point in time or maturity from which to reckon. We need a zero more fundamental than that of birth. We could be content with some epochal zero mark in the foetal or embryonic sector of the life-cycle. Lacking the trigonometry or a clinical criterion for the location of such a proximate prenatal zero, one naturally suggests that the basic ontogenetic zero be placed at conception. This placement is competent to correct in some degree the enormous variability in the period of gestation, which makes the natal zero so seriously inadequate for the systematic study of neonatal behavior. This location of the "absolute" zero can be best justified on the pragmatic grounds that it is needed for the chronological age scale.

THE CALIBRATION OF AGE

Time in the Newtonian sense flows at a constant rate; but development flows with some kind of negative acceleration. There is an involuted compression of development in the early stages of the life-cycle. Although a logarithmic curve exaggerates the simplicity and uniformity of growth, it conforms to this characteristic of deceleration.

---

[2] L. L. Thurstone (29) has discussed the absolute zero in intelligence measurement, from a statistical and psychometric viewpoint. He finds on seven psychological test scales a linear relation between absolute variability and the mean test performance of successive age groups. Absolute scaling is used instead of raw scores. The absolute zero is located by extrapolating this linear relation to the scale value of the mean performance at which variability vanishes. His mental-growth curves locate this point at birth or shortly prior to birth.

This means that in developmental economy the value of any given unit of age is quite relative, and is dependent upon the secular location of that unit in the life cycle. A proportionately vast amount of growth is achieved in the prenatal stage, which is divisible into three fairly defined periods—the germinal, the embryonic, and the foetal. The time sequence of the series of life-periods up to adolescence is approximately as follows:

0 Conception (gametes)
0-1 week —Germinal period (zygote)
1-7 weeks—Embryonic period
7-40 weeks—Foetal period
40-80 weeks—Early infancy (perception-prehension)
80-144 weeks—Later infancy (language-locomotion)
2-6 years—Early childhood
6-12 years—Later childhood (pre-adolescence)

An inspection of this time table will suggest that the younger the organism, the greater is the developmental value of a given unit of time. This holds generally true for physical growth and also for mental growth. For the period of later childhood a measuring scale for mental development may be graduated in terms of years, like the Binet scale; but the maturity of the infant must be reckoned in months and in weeks. Indeed, daily increments of behavior development are not beyond demonstration in human infancy. In careful studies of the infancy of lower animals hours may even come into the reckoning. For example, the systematic study of the foetal behavior of the white rat requires knowledge of the very hour of insemination.

So important are small time intervals in the normative and comparative investigations of human infancy that the solar month has serious disadvantages. The variations in the length of the solar month are a source of error. There is a difference of 11% between the length of February and the length of March, a difference by no means negligible. Accordingly the lunar month as opposed to the solar month is much more satisfactory for the organization of age norms of development.[3] The adoption of the lunar month as a basis of specifying infant age would make the experimental and normative data of different workers more comparable.[4]

---

[3] The Yale Psycho-Clinic has adopted the lunar month as the basis for formulating norms throughout the first two years of life. During the second year it is at present convenient to use longer intervals. This is accomplished by starting a quarterly (12 weeks) reckoning at the age of 56 weeks, making the interval sequence as follows:—52 weeks (1 year); 56 weeks (12.08 mo.); 68 weeks (15.75 mo.); 80 weeks (18.5 mo.); 92 weeks (21.25 mo.); 104 weeks (2 years). This calibration has the advantage of preserving the lunar month sequence intact throughout the second year without deviating too widely from the Gregorian quarterly intervals. It is also a normative advantage to have an additional comparative set of norms at 56 weeks to permit finer specification and measurement of the characteristics of the one-year-old.

[4] The National Academy of Sciences, recognizing the disadvantages of the Gregorian calendar for scientific purposes, has lent its support to the growing international movement, assisted by the League of Nations, in favor of a simplified fixed calendar providing 13 equal months of 28 days each.

THE PATTERNS OF BEHAVIOR

In spite of its inexhaustible complexity and variability, the developmental stream of early human behavior is so ordered by biological laws that it may be systematically studied. There is indeed no such thing as utter randomness even in the behavior of the sprawling infant. The more closely his very sprawling is studied, the more pattern it assumes. Never does a picture of behavior become as amorphous and as meandering as a drifting cloud. The behavior of foetus and of infant is by nature obedient to pattern. Accordingly the term random activity scarcely has a defensible absolute status. Surely the random activity of the four-week-old infant is distinctive from that of an eight-week-old infant. It is distinctive because it has its own pattern.

The term pattern must here be used mainly in a descriptive sense for its serviceableness in the systematic delineation of developmental data. The term pattern can also be used in a manner which makes it approximate the concept of Gestalt or configuration. A systematic, objective type of developmental psychology cannot dispense with an orderly specification of the emergence and succession of patterns of behavior in the former sense. While recognizing the profounder, interpretive aspects of the problem of pattern, it is necessary to characterize the process of behavioral ontogenesis as though it had organic structure. This structure even though it is constantly dissolving in a process of becoming can be envisaged in its momentary patterns. These moments of pattern are countless, but if they are well selected they will furnish an outline of developmental progression.

It does not follow however, that the term *pattern* must be used only analytically for discrete bits or segments of behavior. The concept of pattern is so protean that it may be applied at will to a microscopic interval or area of behavior like the reflex wink, or to the characteristic course of behavior of the whole infant for a given day or hour, or even to the characteristics of career or individuality displayed throughout the whole of infancy. Although this implies an enormous range of data, there can be no question that present-day research in child development and in the related psychobiological sciences is directed toward just such goals.

The word pattern has the merit that it requires the student to think in quantitative terms of time and space and age. From the most rigorous standpoint, no pattern is completely described until it is expressed in these terms.

A comprehensive syllabus of different types of patterns would occupy a whole chapter or volume. The following list summarily classifies the different fields for pattern study:

1) Patterns of specific behavior
2) Correlated patterns of behavior
3) Durational patterns of behavior
4) Patterns of maturity
5) Patterns of genetic sequence
6) Biogenetic patterns

A brief paragraph must suffice to illustrate these different areas for pattern delineation.

1) *Patterns of Specific Behavior.* Any accessible area of behavior may be selected for discrete study, and formulated in terms of time, space, and form,—the tactile reflex wink, the visual reflex wink, the erection of the head, the clasp of a rod, the assumption of erect posture, the drawing of a circle. With the aid of instrumental technique the quest for pattern can be carried to minute corners or phases of a given behavior. There is little likelihood that a technique can be pushed into territory where pattern in some form ceases to be displayed.

2) *Correlated Patterns of Behavior.* Although any single behavior may be analyzed in its own right, different individual patterns are themselves constellated. One form of behavior tends to occur in association with another form of behavior. Eye movements, at first independent, come to occur in relation to head movements; the dynamic interrelation changes with age. Likewise eye movements become associated with hand movements. These dynamic correlations are expressions of pattern.

3) *Durational Patterns of Behavior.* How does the four-weeks infant invest a day? How will he invest a (durationally) similar day at the age of 24 weeks? What does the elder infant do with five blocks in five minutes? What did he do with them at 20 weeks? What will he do with them at 28 weeks? Such questions, in themselves trivial, become significant when approached from the standpoint of behavior economy. The developmental changes in this economy manifest themselves in increments or alterations of clock-unit patterns of performance. Expressed in behavior, duration takes on design. These designs can be experimentally explored.

4) *Patterns of Maturity.* The complex of behavior growth has altitude as well as temporal amplitude. The very use of the terms maturity-level and maturity-stage suggest this aspect of development. These stages of maturity may be studied horizontally as though by cross-section. The cross-sections may be mapped. Such horizontal mapping, whether limited to the individual or to a statistical group, is a scientific effort to formulate underlying pattern.

5) *Patterns of Genetic Sequence.* The complex of behavior growth also has a longitudinal aspect. It is made up of strands or channels projected in a linear manner. The development of perception, of postural control, of prehension, of locomotion, of adaptive exploitation, of language comprehension and expression, of social behavior in general and in special aspects follows continuous lines of genetic sequence. The orderly study of these lines of sequence yields another quota of patterns for the interpretation of infant development.

6) *Biogenetic Patterns.* In a profound organismic manner all the foregoing patterns are consolidated in the individuality of the infant himself. The concatenation of all the patterns of the individual is to be regarded as an expression of pervasive super-pattern. In its most generic aspects this super-pattern expresses the characters of the species, possibly also of the race, and of the constitutional type. Patterns are themselves patterned. Just as the quest for pattern in the descending microscopic

direction is unlimited, so the search for pattern in the opposite direction of ascending individual integration is capable of indefinite elaboration.

The Investigation of Behavior Pattern

The study of different kinds of behavior pattern requires varied methods of approach. Several methods may be briefly specified as follows: (1) neurophysiology; (2) experimental analysis; (3) psychometry; (4) naturalistic observation; (5) clinical study; (6) comparative method; (7) the method of co-twin control.

1) *Neurophysiology.* The fundamentals of the mechanics of behavior and of patterns of performance may be sought in the lower animal forms. This is well illustrated by the notable work of Coghill in his correlated anatomical and physiological studies of the growth of the nervous system in Amphibia. His approach has been at once structural, embryological and functional. The researches have been refined even to the point of cell count delineation of various parts of the nervous system in *Amblystoma Punctatum.* The development of cell structures and cell groupings have been studied in relation to developing function, and in them Coghill has found the prototype of integrating mechanisms in vertebrates. He has described the neural counterparts of behavior patterns in the field of locomotion.

Such psychobiological studies go beyond the ordinary scope of animal and human psychology; but are significant for the interpretation of the developmental aspects of behavior. In the higher animals similar problems have been explored by experimental interference with the nervous system. In man histological study of lesions in the nervous system of the infant may be brought into relation with antecedent studies of behavior.

2) *Experimental Analysis.* Behavior pattern may be studied by instituting controlled situations and by observing or recording response in terms of time and space. Although infancy presents difficulties with respect to the control of conditions, many of the basic experimental procedures used with animals and adults may be adopted.

The cinema is a peculiarly suitable tool for the systematic charting and comparative study of behaviors at ascending developmental stages. To a considerable degree the cinema registers the temporal and spatial values of the behavior pattern. The film consists of a series of frames, ordinarily 16 per second of duration. Each frame therefore records the momentary configuration of the pattern and preserves the attitudes of head, eye, body, hand, fingers, in their dynamic correlation. A consecutive study of successive correlations yields a picture of the behavior event—yields a pattern. By measuring, counting, tracing, and graphing the components of the behavior it is possible to express the patterns in compact objective and also quantitative terms.[5]

3) *Psychometry.* Any test situation which elicits characteristic behavior, by the same token becomes a device for the study of pattern. A

---

[5] A technique for the study and record of such patterns is being worked out in the experimental photographic laboratory of the Yale Psycho-Clinic.

well-conceived, well-administered test is neither more nor less than an experiment. Ideally there can be no rigid distinction between psychometric and experimental methods. Psychometry, however, suggests measurement more than analysis or elucidation.

In principle the methods of mental measurement are as applicable to infancy as to childhood or youth. There is no impassable gulf between drawing a diamond and an infantile banging of a spoon. Both are psychomotor reactions symptomatic of behavior equipment and maturity. Both may be investigated statistically in selected and unselected children and in clinical deviates. Both may be placed, if not in series, in a scale of developmental items, for the "measurement" of maturity.

The application of psychometric criteria must be made with great caution in the period of infancy. It is well nigh impossible to validate test items by purely statistical treatment of data. The impossibility of determining the true postconceptional age of the infant frequently operates as a source of error. Oversimplifying numerical expressions can be avoided by refining and diversifying psychometric techniques in the direction of analytic characterization.

4) *Naturalistic Observation.* This is observation of the natural life of the infant free from the constraints of experimental and test situations. It includes all of the marginal factors which influence experimentally observed response; it may be made to embrace the total behavior day and even the life-career of the infant.

The data from naturalistic observation may easily become diffuse and prolix, but this is not an inherent weakness of the method itself. With the aid of systematic experimental procedure, the naturalistic method can be made critical and selective. In a sense it is more difficult and exacting than experimentation. It is peculiarly necessary in the field of infant behavior, as a corrective to misleading interpretation of experimental results.

5) *Clinical Study.* On its highest level clinical study combines both experimental and naturalistic observations. It focuses on the total individual or on special conditions in relation to the individual as a whole. It is applicable to normal as well as unusual manifestations of behavior; but it is particularly appropriate to those departures from the normal in which nature virtually provides an experiment for elucidation. In this sense the study of the deviating infant lies within the field of experimental etiology, even though the experiment is neither planned nor fully controlled. Harvey's remark, athough made in another context, is pertinent: "Nature is nowhere accustomed more openly to display her secret mysteries than when she shows traces of her workings apart from the beaten path."

Naturalistic and experimental observations are not mutually exclusive. Rather they are reciprocally corrective and are typically combined in a clinical approach.

Behavior symptoms figure in numerous diseases and developmental defects. Such diseases do not only have pathological entity in the anatomical and physiological sense; they have entity in the sphere of behavior. Behavior symptomatology in infants, when systematically studied, becomes a part of clinical medicine.

6) *Comparative Method.* Comparative investigation is a special adaptation of clinical method. It is peculiarly suitable for the study of problems of pattern in their developmental aspects. Patterns cannot be comprehensively measured in the absolute sense. They cannot be arranged in linear continuity along a mensuration scale. Patterns occur in hierarchical orders of complexity. These orders can be graded and they can be compared. Contiguous orders, if the gap between is not too small, throw each other into partial contrast. Judgments can be made as to degrees and qualities of contrast. These judgments may simulate and even approximate absolute measurement; but, strictly speaking, they are relative. It is the relativity of the judgments which makes them fruitful for the study of developmental characteristics. Development as such cannot be measured in absolute units of development. Practically and perhaps epistemologically, such a metric is not at our disposal. But stages of development can always be compared, objectively compared, with the aid of instruments of precision if necessary.

By cinema or by bodily presence a 24-week infant and a 28-week infant may be placed side by side—in dorsal position, prone position, or seated at an examining table, which becomes the stimulus stage for presenting a toy, a pellet, a string, a piece of paper. In all of these situations characteristic patterns of response arise in convenient juxtaposition. The very profiles of the patterns may throw each other into relief. Intricate complexes of attitude and temporal ordering for which we lack a lexicon emerge for relational inspection. One child almost literally becomes a "measure" of the other. Systematically pursued, and channelled by experimental limitations, this comparative method becomes a tool for the genetic delineation of pattern.

7) *The Method of Co-Twin Control.* The method of co-twin control has been described elsewhere in detail, and is proposed as an experimental method for analyzing biogenetic problems such as the relations of growth and learning in infancy. This method may be regarded as an experimental refinement of the comparative method for the conduct of extended developmental investigation. In identical infant twins, nature provides a stage for observing the effects of a developmental stimulus which may be deliberately confined to one twin. By comparative cross-reference the co-twin becomes a control and a check both upon observations and conclusions. The experimentally stimulated twin may be designated as Twin T; and the co-twin reserved as a control, as Twin C. Fraternal twins, although useful for certain types of comparative study do not lend themselves to this kind of experimental control (13).

With the aid of cinema records the scope of the method can be widened. Also with such aid it is possible to establish not only increments of skill but the interrelations between acquired skill and varying maturity levels, by making cinematic comparisons in which the ages of the twins are kept uniform or disparate at the will of the investigator. This utilization of the cinema introduces an interesting and useful kind of relativity into the scheme of developmental study, inasmuch as the subjects of analysis are identical twins.

The reliability of the method is greatly enhanced if the genetic and developmental identity of the twins can be definitely established by physical and developmental examinations. When such correspondence of physical and mental traits is well established, identical twins furnish a unique opportunity for psychobiological study. Comparisons of the behavior of twins have an objective quantitative status even when the absolute units of behavior measurements are altogether unavailable. It may well be insisted that it would be difficult to devise a more comprehensive and delicate instrument of behavior measurement than one twin used experimentally in juxtaposition with an identical co-twin who serves as a standard of reference.

The experimental opportunity and value of the method of co-twin control can frequently be increased by supplementing the first experimental stimulus period of Twin T with a similar deferred experimental period for Twin C. The response of Twin C to this deferred stimulation provides critical data for comparative interpretation.

Observational Arrangements for the Study of Infant Behavior

Although many fruitful forms of observation can be undertaken with extremely simple equipment, the more systematic study of the developmental psychology of the infant requires special arrangements and safeguards. Certain types of investigation should not be undertaken at all, except in a medical setting with ample management and protection. It is scarcely necessary to say that under all conditions hygienic precautions should be taken with regard to the cleanliness of the observational surroundings, sterilization of materials, and avoidance of undue contact with other infants. It is very difficult to generalize briefly details of procedure, because these details vary considerably with the age of the infant. The fatigue factor, for example, varies greatly with the age of the child, to say nothing about individual differences. Fortunately, however, the infant tends to obey the "all-or-none" law with respect to fatigue, and to a remarkable degree displays characteristic behavior without gradual deterioration.

For the sake of concreteness a few of the special observational devices which have been developed at the Yale Psycho-Clinic may be briefly described here, with the assistance of the accompanying photographs.

Clinical Crib (Figure 1, 2). This crib has been devised to standardize developmental test situations. It is built like an ordinary nursery crib, but has many adjustable features which permit the platform and panels to be raised to variable heights. The side rails support a movable table top on which the test materials are put with careful regard for placement and details of presentation. When not in use, these materials are concealed in the pockets of the container bag hung at the rear crib end. This bag is removable and facilitates sterilization of the materials.

Infant Supportive Chair (Figure 3). This chair, built along Morris lines, is provided with a removable canvas back and seat and abdominal belt; it supplies comfortable support for infants with imperfect sedentary

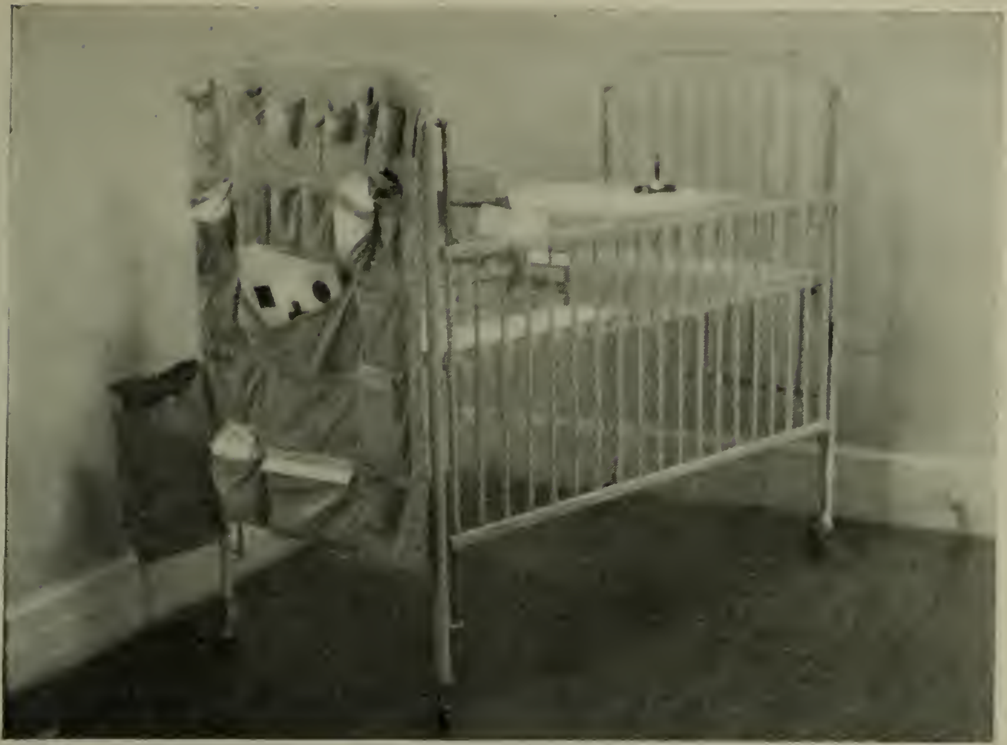


FIGURE 1
Clinical Crib

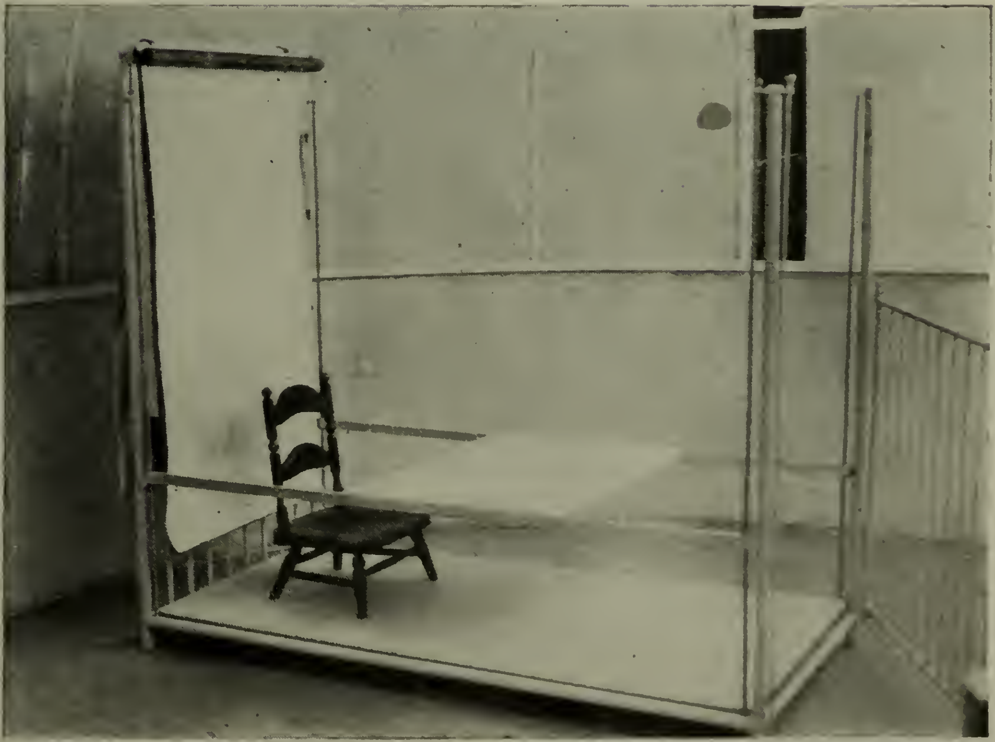


FIGURE 2
Clinical Crib Adjusted for Development Examination

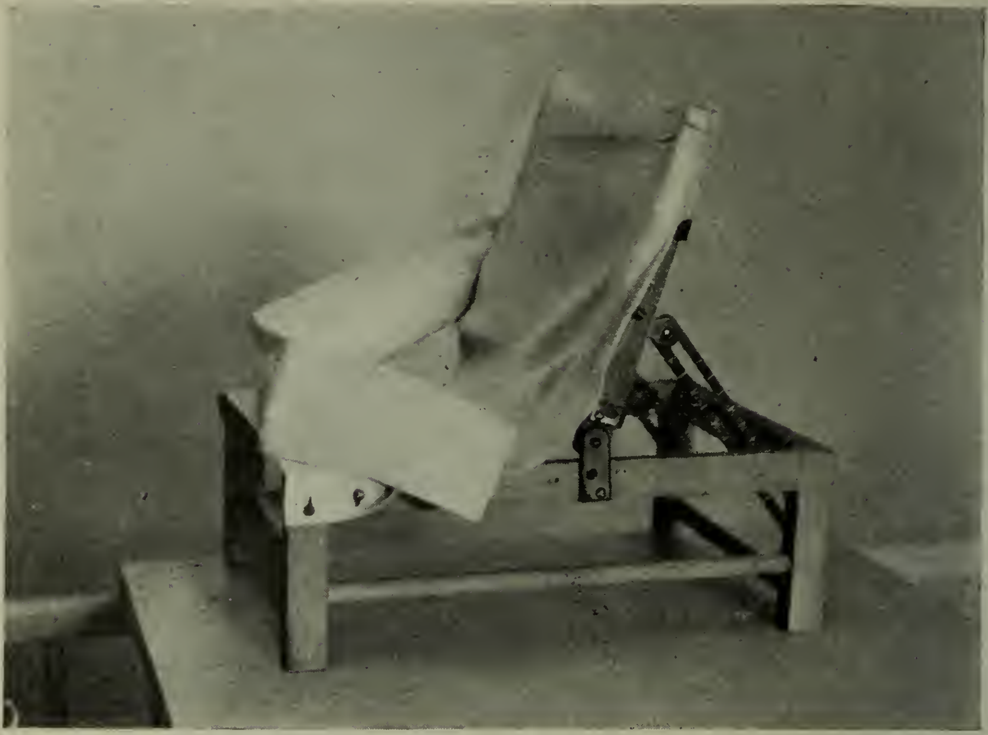


FIGURE 3
Infant Supportive Chair

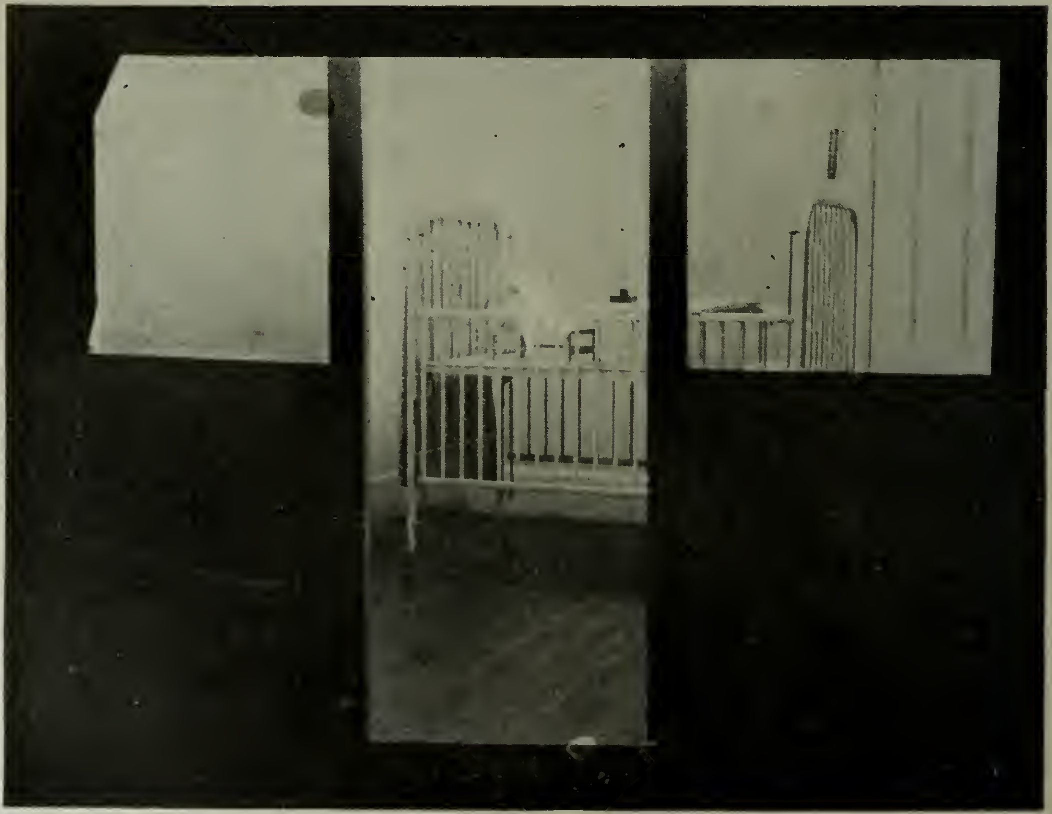


FIGURE 4
One-Way Vision Screen

postural control. Its sanitary and adjustable features make it serviceable.

One-Way Vision Screen (Figure 4). To secure control of infant-behavior observation, it is necessary to remove the distraction of complicating adults, whether parents, recorder, or other assistants. And yet we wish to permit the mother to view the examination. This difficulty has been effectively met at the Yale Psycho-Clinic by a specially devised screen which conceals the observer but not the observed. A folding-panel arrangement of this screen is partly pictured in the illustration. In the actual set-up the screen is a full-length partition, which separates the observers who are stationed in a darkened alcove. The illuminated surface of the screen is painted white, with highly reflective enamel; the dark surface is unpainted. The screen functions as a visual sieve permitting one-way vision only—a useful cloak of invisibility.

The Photographic Observation Dome (Figure 5). The hemispherical structure, 12 feet in diameter, is located in the photographic laboratory

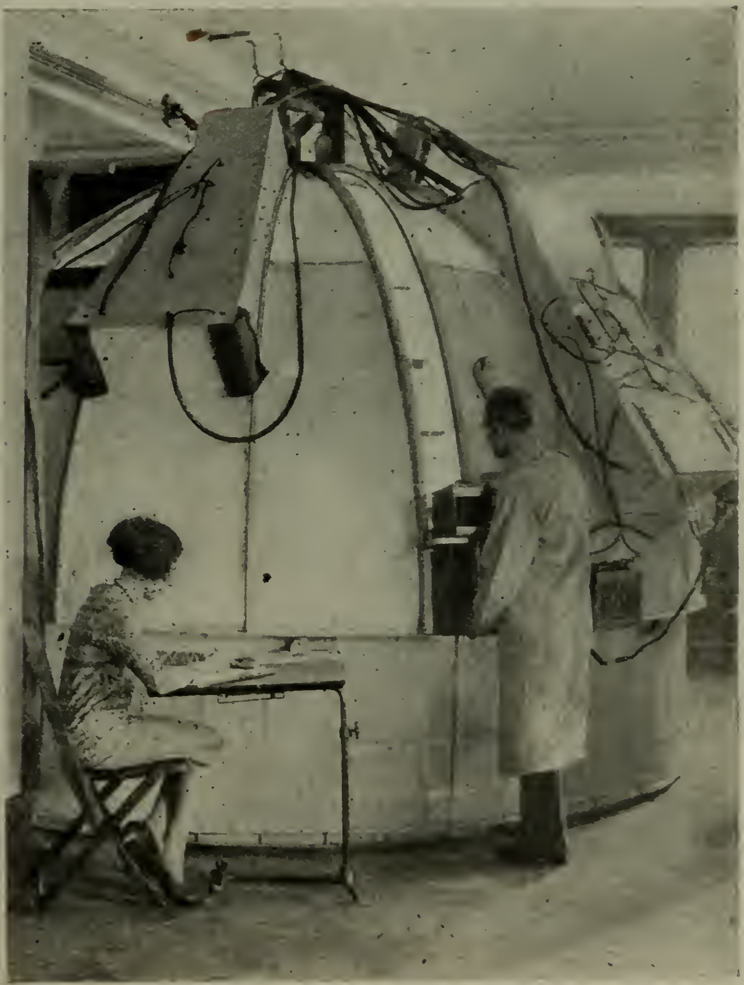


FIGURE 5
Photographic Observation Dome

of the clinic. It is in a sense the reverse of its astronomic prototype. The route of observation is inward, the operator, observers, and parent being stationed without, segregated by the one-way vision screen which encases the dome. The interior is illuminated with a soft artificial light, so suffused that it offers no distraction to the infant in the clinical crib within.

Instantaneous cameras under electrical control and a motion-picture camera are poised on two quadrants at right angles to each other. The cameras may be placed in any position from horizon to zenith, and are operated by electric switch, singly or simultaneously. Simultaneous pictures give control of the third dimension; direct overhead pictures yield valuable configurational data; the cinema records at once the spatial and temporal patterns of behavior (12, pp. 56-81; 15).

THE RANGE OF INFANT BEHAVIOR

Within the limitations of a single chapter major stress has been laid upon the methods and principles which underlie the developmental study of infant psychology. The chief interest and the significance of infant behavior lie in phenomena of change, of emergence, increment, and decrement. The infant himself alters so rapidly and so diversely that it is difficult to make many generalizations about him. We cannot even take a week to make a thoroughgoing investigation of an individual infant, because by the time the week is over the infant we began to investigate has changed into a being so different that from an experimental standpoint he is no longer the selfsame subject.

So swift and continuous are these changes that in the first six years of life the individual traverses far more developmental ground than he will ever again compass in a similar interval. If there were some volumetric unit for expressing the quantity or current flow of development from birth to the age of 24 years, we should find the resultant graph rising with a rocketing sweep in the pre-school years. If there were some graphic method for expressing the multiplicity and the variety of behavior patterns, the diagram whould show a widely disproportionate range of behavior in the period of infancy. When the wealth of behavior equipment acquired in the prenatal period is cast into the reckoning, the enormous concentration of development in early life becomes yet more striking (see Figure 6).

The new-born babe is almost a pulp from the standpoint of postural control; at two years he can run; by six years he may be acquiring the rudiments of golf or musical skill. In the field of perception he advances from a stage of sketchy, wavering, ocular responses to fine discriminations of color, size, form. At birth he reflexively clasps a rod (or crayon) with eyes crudely wandering or vacantly transfixed; at six he adaptively scans the perimeter of a square or a triangle, reproducing each form with directed crayon. The birth cry, scant in modulation and social meaning, marks the low level of language, which in two years rises to sentence structures, and in six years to elaborated syntactic speech, with questions

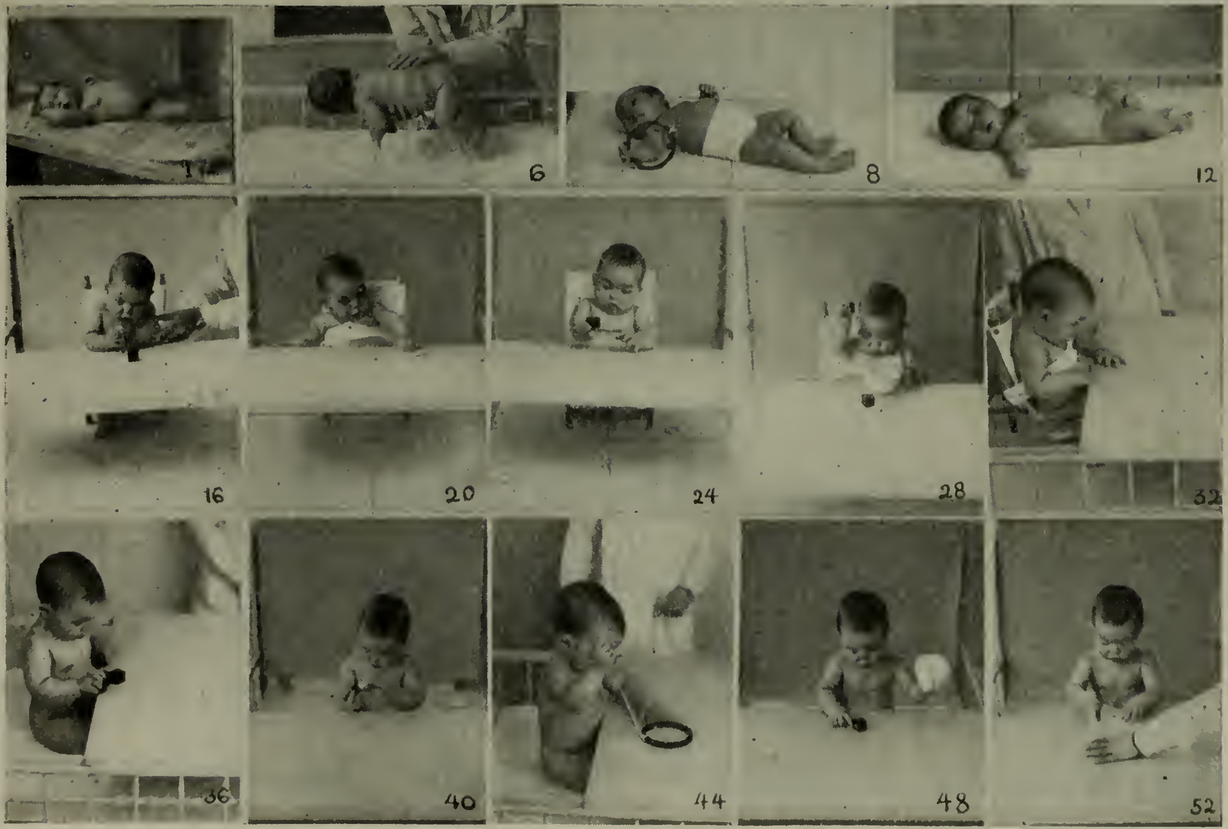


FIGURE 6

Behavior Patterns in the First Year of Life

1 week—tonic neck reflex posture; 6 weeks—head extension in ventral suspension; 8 weeks—reflexive clasp of ring without regard; 12 weeks—eyes follow small ball (note persistence of tonic neck reflex posture at both 8 and 12 weeks); 16 weeks—regards cube; 20 weeks—approaches spoon; 24 weeks—prehends cube; 28 weeks—bangs cube; 32 weeks—pokes pellet with partial thumb opposition; 36 weeks—exploits third cube; 40 weeks—brief inspection of bell clapper; 44 weeks—pulls string to secure ring; 48 weeks—removes cup to secure cube; 52 weeks—makes crude scribble.

and even primitive ideas of causality. In personality make-up there is enormous progress; the school beginner is already so highly organized (both socially and biologically) that he definitely foreshadows the sort of individual he will be in mature years.

The task of developmental psychology is to give systematic formulation to this prodigious complex of early behavioral growth. Inventory is the basis of such formulation, but the very wealth of the data will make increasing demand upon principles of selection which will coordinate and simplify the data. Here experimental procedures have an important scope. In the following sections experimental studies will be sketchily mentioned to illustrate significant lines of investigation in the major fields of behavior.

Studies of Prenatal Behavior

The older interest in the prenatal period centered about questions of maternal impressions and the consciousness of the foetus. Current interest is concerned with the developmental aspects of the problem, which has been experimentally attacked in animals. Studies of the foetal movements of the white rat and the guinea pig have shown that these movements tend in a significant way to follow a characteristic time schedule. The progression is in general from head to foot. The gestation period of the white rat is 480 hours. A definite lateral head movement has been observed at the 378th hour. The behavior development is so rapid that foetal subjects differing by only one hour of age may be planned for in the experimental series.

Although systematic observation of the human foetus is impossible, Preyer (26), Kussmaul, and others have supplied scattered data. Recently Minkowski has reported numerous observations on the human foetus surgically removed by Caesarian section. The embryonic human heart is active as early as the third week of gestation. Slow movements of arms and legs were observed in an embryo of seven weeks. At two months, Minkowski found irradiating reflexes following a light touch of the skin; at three months, mouth and chin movements occurred on tongue stimulation. From the fourth to the fifth month patterned reflexes appear. Stimulating a foot evoked a diagonal (trot) reaction of the opposite hand. Turning the head to one side caused movement in the arm of that side. Bilateral symmetrical movements of arms and legs occurred when the position of the foetus was changed from horizontal to vertical, a complicated labyrinthine reflex originating in the simicircular canals. Plantar flexion of the toes on foot stimulation was also noted at the fifth prenatal month.

Even by the sixth month of gestation the behavior equipment is so well advanced that there is a bare chance of survival of the premature born at that time. Survival after the seventh (solar) gestation month is relatively common. It is as though nature hastened to bring the most vital functions of postnatal life to partial completion against the contingency of premature birth. Postmatures, paradoxically enough, continue their normally postnatal maturation in utero. We have studied the mental-growth curves of both pre- and postmatures and have found that the trend and tempo of development are not grossly altered by the unusual circumstances of birth, if there has been no natal damage.

Such damage frequently occurs. Capper has found, when a large series of cases is considered, that the immature neonate (birth weight less than 2500 g.) and the premature frequently prove to be backward, defective, or unstable in later life.

STUDIES OF PERCEPTUAL BEHAVIOR

Observations of the perceptual responses of premature infants prove that the basic equipment for sensorimotor reactions is well advanced in the prenatal period. Peterson and Rainey found sensitiveness to light, and occasionally to sound in their prematures. Kussmaul reported one instance of ocular pursuit in a presumable premature. Sherman and Sherman found sluggish pupillary contraction at an average age of 7.7 hours in a group of nine new-borns. The pupillary reflex is well established by the second day, suggesting either very rapid growth or the disappearance of complicating physiological impediments. The new-born may prefer darkness and twilight but reacts positively to light even on the first day. This early orientation, Watson regards as "a tropism-like response," different from "fixation" which comes later. Reports concerning coordinated ability to follow moving objects vary considerably. Photographic studies by McGinnis have shown the presence of nystagmus and ocular pursuit in the very young new-born.

The criterion for fixation is not clear. Shinn thought that until nearly three months her niece had not really "looked *at* any special objects which were not lighted up, colored, or in motion." M. C. Jones (17), in a study based on over 300 subjects, showed that horizontal eye-following movements apeared before vertical and the vertical before the circular. Half of her subjects followed a light horizontally by the 58th day; all by the 90th day. Ocular pursuit of rapid movements comes appreciably later and by slow stages. Questing regard for objects after disappearance also matures somewhat gradually. It is frequently observed as early as 32 weeks. Nearly all children look after a fallen spoon at 40 weeks. Asymmetry in the coordination of eye movements is frequent in the first month, and occurs as late as the third and fourth months. Blinking on a tactile stimulus is present at birth. Blinking on the threat of a visual stimulus occurs in most children by 16 weeks.

The coordination of eye and head movements passes through developmental stages which have not been carefully studied. Little is known about auditory localization in infancy. The head turns promptly to a source of sound as early as 16 weeks. At a later age the eyes turn independently.

Tudor-Hart and Hetzer have made interesting quantitative studies of early audition. They found a rapid increase of efficiency in hearing in the first four weeks. Judging by reactions to noises, sounds, and the human voice, the four- to seven-day child heard one and a half times as well as the zero- to three-day child, and the two-month child heard two and a half times as well. Not much improvement was discovered after the second month.

Studies of Motor Development

Motor development furnishes the most concrete field for observation and has been the subject of numerous experimental studies, particularly in the first postnatal month and year. The more fundamental studies have been neurological in approach, and have concerned the integrative functions of the nervous system and the postural reflexes (Magnus, deKleyn, Peiper). Specific reflex responses to various positions (dorsal and ventral and diagonal suspension) have been described by Peiper (23) and compared with spontaneous movements. Burnside (7), with the aid of motion pictures, has made a careful analysis of the coordination of limbs and body postures and the sequence of movements involved in the development of progression, with special reference to walking in nine children from 7 to 18 months.

Buford Johnson has inquired into the relations of muscular control to chronological age, weight-height index, practice and sex, using grip, speed, steadiness, target, and tracing tests. Tracing tests showed consistent age differentiation up to seven years. Gesell has outlined the growth of drawing (crayon and paper behavior) from early infancy throughout the pre-school years, showing a progression in different strokes and forms. Stages in the development of prehension and other eye-hand behavior have been similarly outlined (Figure 7).

M. C. Jones (17) has studied the development of thumb opposition and reaching in 365 infants. These functions are nearly parallel, beginning at about 80 days, attaining a median at about five months. Lippman has studied the differentiating reactions to a first, second, and third object in 178 infants.

Downey has attempted to determine types of dextrality and eyedness in pre-school children, using motor tests, unilateral eye opening and eye closure and sighting. H. E. Jones has determined dextrality ratios by recording the preferential hand use of nursery-school children in peg- and formboard situations. Baldwin made a kymographic study of individual differences in rhythm, as shown by ability to mark time.

The investigation of oculomotor reactions have been briefly mentioned in connection with perceptual behavior. Such investigation depends upon methods of photographic registration and cinematographic recording and analysis. From the standpoint of pattern, these problems lie also in the field of motor development.

Studies of Language Behavior

Language behavior, in its expressional aspects, constitutes an inexhaustible field for developmental study. The literature lists numerous studies based chiefly on naturalistic observation, periodic and diary records, and vocabulary inventories. In the aggregate these studies contain a

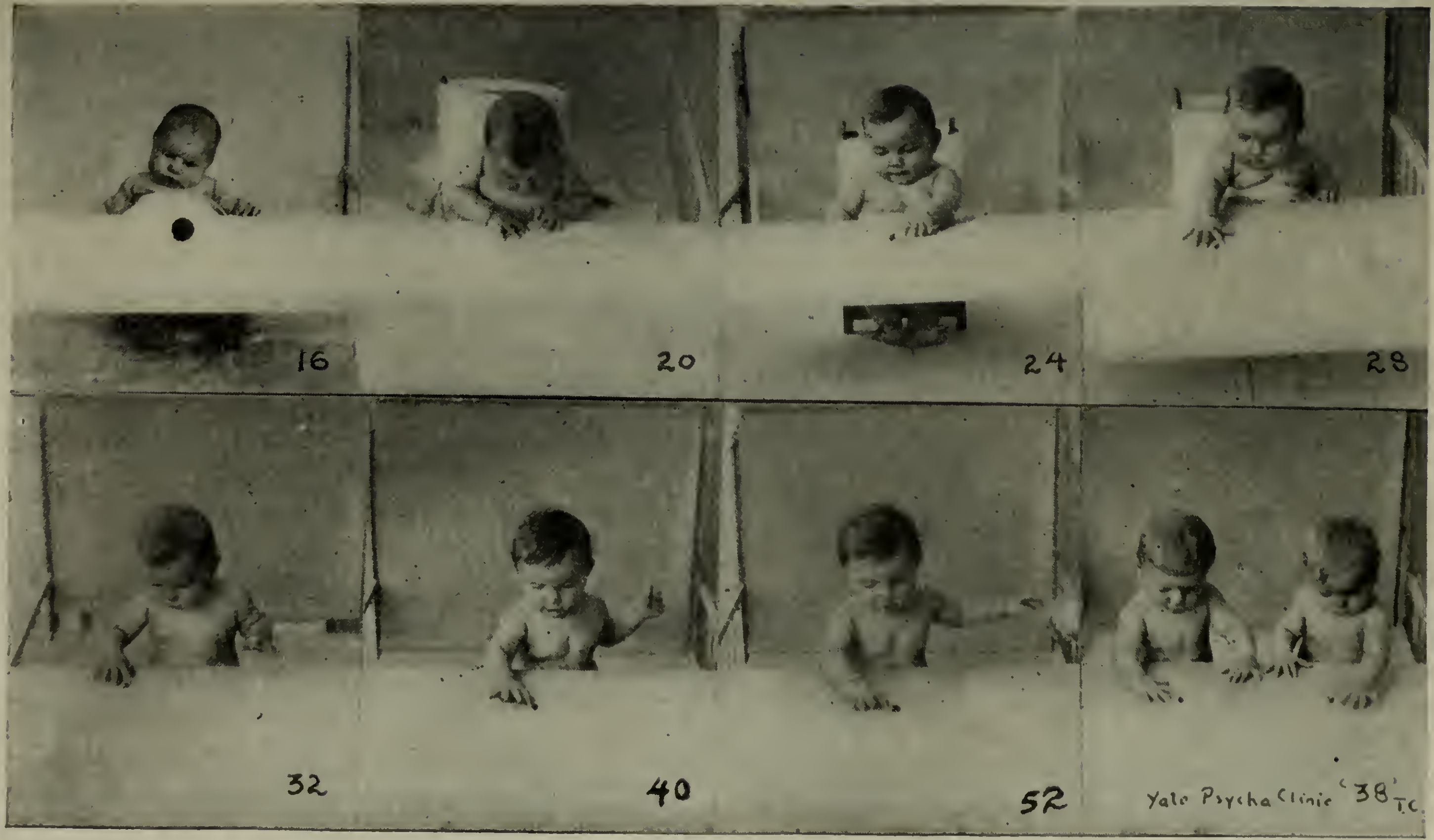


FIGURE 7

Typical Reactions to a Pellet

16 weeks—no visual regard for small pellet, but fixation on a one-inch sphere; 20 weeks—more protracted visual fixation on the pellet with slight postural changes; 24 weeks—unilateral, pronated approach with scratching in the vicinity of the pellet; 28 weeks—pronated hand approach with raking flexion resulting in occasional palmar prehension; 32 weeks—pronated hand approach with extension of index finger and partial suppression of other digits, resulting in poking or prehension by thumb in partial opposition to index finger; 40 weeks—rotation of wrist with pincer-like prehension by index finger and thumb; 52 weeks—perfection and further delimitation of pincer-like prehension; Twins T and C, age 38 weeks, approaching pellet.

wealth of material, but precise experimental investigation has been relatively infrequent. Test methods figure prominently.

Gerbach (10) assembled 19 vocabulary studies made up to the year 1917, showing a range of six words at eight months to 6837 words at five years. Boyd, Gesell, and Stern have reported observations on the beginnings of syntactical speech, the use of pronouns and prepositions. Nice has made studies of speech in relation to handedness. Orton is working on similar problems in older children through an investigation of reading disabilities and speech defects. His theory in regard to the importance of the cerebral dominance of one hemisphere in the normal functioning of speech mechanisms raises significant developmental problems for experimental study.

H. M. Johnson (16) reports in some detail the spontaneous conversations of children under three years in solitary and in nursery-school situations. Piaget (24, p. 75) has made an illuminating analysis of the thought processes of the child in relation to language, showing the highly egocentric aspect of child logic. His method of investigation was essentially clinical, utilizing a naturalistic form of experimentation, designed to test ideas of causality, judgment, and reasoning. His clinical approach is similar to the free-association method for he utilizes or creates a situation in which he lets the child talk, noticing the spontaneous unfoldment of thought. Baldwin made a study of time and number concepts in preschool children (1).

Descoeudres (9) and others have made normative investigations preliminary to the standardization of a scale for the measurement of the language development of the child from two to seven years.

Studies of Adaptive and Learning Behavior

The vast field of adaptive behavior has been approached from two somewhat different directions: that of experimental and theoretical analysis; that of developmental measurement and description. Fundamentally these two methods of approach are one and supplement each other. For example, a limited number of infants may be individually studied to ascertain how they will use a string to secure an attached object. Their reactions may be carefully analyzed to determine evidences of "insight"; but an understanding of the different kinds or degrees of "insight" can be arrived at only by developmental specifications of the different patterns of reaction to the selfsame situation at advancing age levels. Such normative seriation brings out the progressive and emergent aspects of adaptive functions, and makes possible the diagnostic and predictive study of mental abilities and mental defects in early infancy. A broad term like adaptive behavior is necessary to extend the genesis of intelligence down to the earliest ages.

Binet's work (2) and outlook have proved pregnant in this field. Kuhlmann (19), Yerkes (32), Baldwin (1), Stutsman (28), Goodenough (14), Charlotte Bühler (4, 5), Hetzer and Wolf, Figurin and Denisoff, Gesell, and others have investigated the stages of mental development in young children. Some of these studies are predicated on a concept of a general measurable intelligence; other studies are more definitely directed toward the establishment of clinical criteria for the delineation and expression of patterns and levels of maturity.

Koffka (18) discusses the growth of the mind from the more general standpoints of genetic psychology and of Gestalt. Both he and Bühler (6) give much recognition to the explanatory value of Köhler's work in the mentality of apes. Köhler, Kohs, Hunter, and others have somewhat directly carried over methods of animal experimentation into the field of infant behavior.

These methods will probably be increasingly used to determine the characteristics and norms of infant learning, a field which has received surprisingly little attention. Krasnogorski used the method of the conditioned reflex to investigate learning capacities in the normal infant. Mateer (21) standardized a similar method (a chocolate, the adequate stimulus; an eye bandage, the substituted stimulus), and applied it somewhat extensively to 50 normal and 7 defective children of varying ages.

Studies of Emotion and Personality

Experimental investigation of emotional behavior and of personality in young children has been meager. The general literature itself is extensive, as is shown by the fact that the National Research Council Bibliography (Circular No. 72) lists over 1300 titles (up to the year 1926) on "The Analysis and Measurement of Human Personality." But only a few of these titles deal with the period of infancy or with experimental method.

The method of conditioning and reconditioning emotions was used by Watson in his well-known studies of fear. M. C. Jones has used similar methods in a study of the elimination of fears in 70 orphanage children from three months to seven years of age. Gates investigated the growth of social perceptions by classifying the responses of children from 3 to 14 years old, to pictures portraying joy, anger, fear, scorn, etc. Charlotte Bühler observed the earliest social interreactions of young children (4 months to 22 months) by placing them together in pairs with common toys and in other similar situations. R. W. Washburn has made a developmental survey of smiling and laughter by means of periodic studies of 15 children at lunar-month intervals throughout the first year of life. A standard set of stimuli was used, and when the reactions and patterns were studied four expressional types of personality emerged. These types remained consistent.

L. R. Marston contrived experimental conditions to observe the emotional reactions of 100 children (two to six years of age) with respect to introversion and extroversion. The criteria were social resistance, compliance with the request of the experimenter, interest in novel situations, and self-assertion.

Psychoanalytic literature contains material on the psychogenetic significance of early emotional experience. Leonard's autobiographical volume, *The Locomotive God,* though literary in construction, contains a remarkable account of an adult phobia traced in detail to a fixation in infancy.

A significant psychobiological approach to the problem of personality is represented in Dr. Walter Jaensch's clinical studies of the T-type (tetanoid) and B-type (Basedow). These types have been elaborately studied and differentiated with reference to excitability to mechanical and galvanic stimulation, optical characteristics including eidetic imagery, emotional tendencies, reaction to calcium therapy, physiognomical and mimetic traits, and finally the microscopic structure of the genetic strata of the capillary system of the skin. This correlation of the psychophysical aspects of personality is indicative of the modern biological trend in studies of individuals and of types.

The Maturational Basis of Behavior Pattern

The influence of conditioning on the human infant has been forcibly asserted from the staindpoint of behaviorism. Emotional organization, handedness, and even the reflex visual wink have been interpreted as conditioned responses. The effects of conditioning have been so frequently demonstrated and discussed that it is desirable to examine the influence of sheer maturation on the behavior patterns of infancy.

The development of the prematurely born infant throws significant light on the general problem. Our clinical studies of both healthy premature and postmature infants have repeatedly shown that the growth of their behavior tends to be obedient to the regular underlying pattern of genetic sequence, irrespective of the irregularity of the birth event.

Refined studies will doubtless reveal that such irregularity does subtly modify many details of behavior; but as a point of departure for the discussion of maturation, nothing is more comprehensive in implication than the general stability of the trend and the tempo of development, in spite of precocious or postponed displacement of birth. The patterns of genetic sequence insure a basically similar growth career for full term, pre-term, and post-term infants. It is as though nature had provided a regulatory factor of safety against the stress of extreme variations of environment. In the mechanisms of maturation this regulation operates.

The term growth may be construed to embrace the total complex of ontogenetic development. Maturation refers to those phases and products of growth which are wholly or chiefly due to innate and endogenous factors. It is our purpose to assemble, in a summary manner, diverse evidences of behavior maturation, based upon our clinical, experimental, and normative observations.

These evidences are drawn from several sources as follows:

Prehension Patterns and Maturation
Maturational Correspondence in Identical Twins
Maturation and Learning in Twins
Maturation and Physical Handicap
The Maturational Aspect of Emotional Behavior

Prehension Patterns and Maturation

The development of prehension throughout the first year of life displays significant progressive changes in behavior pattern. These changes raise searching doubts concerning the influence of experience and training upon these patterns. We have studied these changes with particular reference to a pellet eight millimeters in diameter. The characteristic eye-hand reactions of an infant confronted with this tiny pellet may be recapitulated in the following, which is a genetic order:

a) No visual regard for the pellet
b) Transient regard for the pellet
c) More prolonged and definite fixation upon the pellet with slight postural changes
d) Visual fixation with crude bilateral or unilateral hand approach. (20 weeks)
e) Unilateral pronated hand approach with scratching in vicinity of the pellet. (24 weeks)
f) Pronated hand approach with occasional raking flexion resulting in palmar prehension. (28) weeks)
g) Pronated hand approach with extension of index finger and partial suppression of other digits, resulting in poking or prehension by index finger with partial thumb opposition
h) Rotation of wrist in hand approach, with pincer-like prehension of pellet by index finger and thumb. (40 weeks)
i) Perfection and further delimitation of pincer-like response

All these changes mature with subtle but significant correlated changes in head posture, body posture, hand and arm attitude, and associated visual behavior. It seems quite erroneous to say that the child learns to prehend the pellet in the traditional sense of the learning process. Crudely, but nevertheless effectively, he prehends the pellet by gross palmar approach as early as the age of 28 weeks. The refinement of his eye-hand behavior comes not by the alleged utilization of snatches of successful random activity but by the progressive acquisition and consolidation of a hierarchy of behavior patterns which are the result of developmental decrements and increments rather than the stamping in or chaining of satisfying, successful reflexes. The defective child shows retardation in the acquisition of these patterns even though he may, in a durational sense, have a larger fund of prehensory experience. It is not improbable that many of these developmental changes in the pattern of prehension would be realized even if the prehensory hand were altogether swaddled and deprived of activity. When the prehensory mechanism is damaged by restricted birth injury to the brain, resulting in extensive athetosis, the propensity to prehend or reach may still assert itself at the proper genetic level. Even though the propensity is aborted its presence is highly suggestive of the potency of maturational determination.

Maturational Correspondence in Identical Twins

In a recent study we have gathered extensive detailed data on the development of prehension in a pair of identical infant twins. These twins were identical not only with regard to their skin patterns but also, to a remarkable degree, with regard to their behavior patterns. Nowhere was this more objectively shown than in their prehensory reactions to cubes and pellets under controlled observational conditions. At 28 weeks both of these twins, being somewhat retarded in their development, were visually unheedful of the pellet, though they definitely regarded a cube. At 38 weeks they addressed themselves in an identical manner to the pellet. The hands were in full pronation, the fingers, spread apart in a fan-like manner, were fully extended. The thumb was fully extended almost at right angles. The photographic record of their attack upon the pellet, in the motion pictures, shows an almost uncanny degree of identity in the details of postural attitude, hand attitude, and mechanism of grasp. (See Figure 7.)

At 40 weeks each twin made a crude raking attack upon the pellet, with occasional awkward but completed prehension in which the palm and all of the digits participated. The form of the prehension pattern was again remarkably similar in both children. At 42 weeks they were again examined in the same situation. Although there had been no special instruction or conditioning in the interval, these two weeks imposed a palpable and strikingly similar change upon the prehension picture. Simultaneous flexion of the digits was very neatly displaced by a preferential flexion of the index finger. The raking approach was replaced by a poking with the tip of the index finger. Such an interesting inflection of the prehensory pattern surely could not have been induced so precisely and so simultaneously in both of these children without the presence of controlling factors of organic maturation. Of similar significance is the fact that comparable changes in prehension pattern appeared coincidentally throughout the course of their development.

The correspondences in behavior patterns in these twins were literally uncountable. However, the records of 13 developmental examinations were analyzed, and 612 separate comparative ratings of behavior items were made from these records in order to determine items of correspondence and disparity. There were 99 items of minor disparity and 513 items of identical or nearly identical correspondence. The parity of behavior patterns was overwhelming.

Many convincing examples of behavior correspondence might be cited. We content ourselves with a few, much-abbreviated illustrations. Here is one which seems to us to have experimental control, even though it deals with nothing more that the reaction of two infants when placed in exactly the same manner upon a flat platform to observe their postural control in the sitting position. Both children showed precisely the same kind and degree of difficulty in equilibrium at the age of 28 weeks. In both there was a tendency to sway to the right; in both it was impossible, even by spreading the legs, to make the body lean forward sufficiently to establish a passive balance. In the case of each child there was an antagonistic tension which made the child rebound backward in an automatic manner resembling a sharp spring-like action of a knife blade snapping into position. We have never seen precisely this kind of reaction in an infant at this age. It is inconceivable that the response arose out of some identical conditioning factor in the environment. It is reasonable to suppose that this distinctive behavior pattern reflected a maturity level and a synchronous neural organization shared by both children because of their common genetic origin.

Within a week this reaction disappeared. Four days later the twins were placed upon a large blotter on the platform of a clinical crib and maintained the sitting position by leaning forward. Simultaneously they attacked the blotter with the hand in full pronation, and simultaneously with vocalizaton they continued to scratch the blotter, leaving visible marks. Here again was a dramatic bit of correspondence all the more impressive because displayed simultaneously. The complexity and nature of these two behavior patterns again suggest the determining rôle of maturation. If it is argued that extrinsic factors determine the form and the time incidence of these simultaneous patterns, it is necessary to demonstrate in detail the cunning arrangements of environment and of conditioning stimuli which could design so precisely, and in duplicate, the configuration of behavior. How can the environment, even of twins, accomplish such architectonic miracles?

A brief example of behavior correspondence may be cited from the 44-weeks examination record. The twins were confronted with a test performance box with its three holes. The common method of approach of the two children, their preferred regard for the edge of the performance box, the fleeting regard for the holes, the exploitation of the vertical surface of the performance box by a scratching, simultaneous flexion of the digits, the failure to place a round rod into any of the holes, the brushing of the surface of the performance box with the rod, the transfer of the rod from one hand to the other, and finally an almost simultaneous peculiar clicking vocalization in both twins—altogether constituted a very complicated behavior pattern, but one which bristled with numerous identities of spatial and dynamic detail. One can give due weight to the significance of this correspondence only by reflecting on the myriad of behavior exploitations of the situation which the twins *might* have adopted. But in spite of this multitude of exploitational possibilities, the twins were apparently under common inner compulsion to adopt those very similarities of behavior which have been noted.

Still another and very pretty example of identity was disclosed in the pellet and bottle test at 48 weeks. This test involved a bit of learning as well as perception and prehension. Three trials were made with each child. The examiner dropped a pellet into a small glass bottle and then gave the bottle to the child. Both children watched the dropping of the pellet with the same transfixed attention. Both children, on the first trial and on the second trial, too, seized the bottle apparently heedless of the contained pellet. Both children on the third trial pursued the pellet by poking at it against the glass. Here the details of behavior pattern extended even into the marginal zone of adaptation through learning.

In passing it should be noted that, although these observations on twins are comparative, they are objective, and may claim the objective quantitative validity which was referred to in a discussion of the method of co-twin control.

Maturation and Learning in Twins

While the positive results of training and conditioning have somewhat obscured the factors of maturation, the limitations of training may be adduced to show the existence of these factors. Such limitations were put to experimental study (13) in the same pair of twins (Twin T and Twin C) whom we have just cited. At the age of 46 weeks, when the thoroughgoing mental and physical identity of the twins had been well established, it was decided to determine the influence of training confined to one twin, by using an experimental method which we have designated the "method of co-twin control." T became the trained twin; C was reserved as a control.

Accordingly, Twin T was systematically trained for 20 minutes daily over a period of six weeks, in two fields of behavior—stair-climbing and cube behavior, including prehension, manipulation, and constructive play with a dozen one-inch red cubes. An experimental staircase arrangement of five treads was used, and for 10 minutes daily Twin T was put through her paces. At 48 weeks she scaled the stairs for the first time with slight assistance. At the conclusion of the six-weeks training period (age, one year) she was a relatively expert climber. At that age her untrained co-twin C would not yet scale the staircase, even with assistance. At the age of 53 weeks, however, when C was again confronted with the staircase she climbed to the top without any assistance and without any previous training whatsoever. In this sense the form and the efficiency of her pattern of climbing were almost purely a function of the maturation of the appropriate neural counterparts.

Twin C was then given an experimental course of training in stair-climbing, two weeks in length. At the end of this period (age, 55 weeks) she approached Twin T in her climbing skill. By means of the motion picture it was possible to make a comparison of the climbing ability of C at 55 weeks (after two weeks of training) with that of T at 52 weeks (after six weeks of training.) This comparison introduced an interesting form of relativity into the investigation and brought out the significant fact that, although T had been trained three times longer and seven weeks earlier, this advantage was more than overcome by the three weeks of C's added age. Again the powerful influence of maturation on infant behavior pattern is made clear. Early training altered slightly the form of the pattern and hastened the acquisition of facility, but left no considerable or decisive locomotor advantage in favor of Twin T.

In the field of cube play the experiment clearly showed that training had no significant effects upon the patterns of prehension, manipulation, and constructive exploitation. Although Twin C had enjoyed no special opportunities in the handling of cubes, her cube behavior was fully equal to that of T after a six-weeks training period. The similarity in temporal and spatial details of pattern was confirmed in this case by a time-space analysis of the behavior patterns by means of the cinema record. This does not, however, mean that there were no changes in the patterns of cube behavior during the training period from 46 weeks to 52 weeks. On the contrary, the records, when analyzed, show consistent and incontrovertible weekly increments. Indeed, a day-by-day analysis of the diurnal records of cube behavior satisfied us that there was a daily drift toward progressive changes in the cube performance patterns. These changes were developmentally achieved by steady processes of decrement and increment rather than by a saltatory or zigzag course. There may be spurts and plateaus and rhythms in the development of other fields of behavior, but at this stage of the life-cycle there was a relatively constant trend toward daily change. This progressive daily changing apparently occurs by a process of continuous emergence which tends to lift the level of development slowly and steadily as though by tide action rather than by rhythmic spurt. We would explain the resistance of the patterns of cube behavior to the influences of training and conditioning by the fact that these patterns are basically under the stress and the regulation of the intrinsic organic factors of maturation. The very fact that there is a growth trend toward daily change of pattern makes the behavior less susceptible to stereotypy and to conditioning.

Maturation and Physical Handicap

This subject opens up the vast field of experimental etiology in which the results of disease and environmental abnormality may be analyzed to determine the influence of extrinsic factors upon the complex of growth. In many instances these extrinsic factors seem to be much less powerful than one might suppose. Even grave degrees of malnutrition, correlated with excessive subnormality of weight, are usually incompetent to inflict any drastic changes upon the forms of fundamental behavior patterns and upon the genetic order of their sequence. While it must be granted that certain food deficiencies, for example, in the field of calcium metabolism, may definitely influence the general picture of behavior, the nervous system itself is remarkably resistant to general adversity, even to malnutrition. When certain areas of the nervous system are actually damaged by disease or injury, maturation cannot make amends, but the maturation of the nervous system seems to proceed toward the optimum in the areas unimpaired, even though lacking the stimulus of exercise of the functions controlled by the impaired areas. It is for this reason that various clinical types of profound motor disability attain none the less considerable approximation to normality in certain patterns of behavior.

In this context we may also mention the high degree of autonomy which the nervous system maintains even in extreme cases of *puberty praecox.*

We have investigated one case in which there was a precocious displacement of puberty amounting to a whole decade. This girl became physiologically mature at the age of 3½ years. In spite of this extreme developmental alteration, the course of her behavior development in the fields of intelligence, language, and locomotion has been relatively normal and stable up to her present age of 11 years.

Here, also, should be mentioned again the general developmental course of the healthy infant born after an abnormally short or an abnormally long gestation period. A premature postnatal environment and a protracted uterine environment must be considered as drastic deviations from normal environmental influence. The relative immunity of the behavior patterns from these environmental deviations again bespeaks the potency of maturational factors.

THE MATURATIONAL ASPECT OF EMOTIONAL BEHAVIOR

The rôle of maturation in the control of emotional behavior has had scant recognition. The primary emotions have been discussed as though they were elementary stable phenomena subject only to the changes of social conditioning. This is the implication in much that has been written concerning the emotion of fear. It seems to us that the problem has been oversimplified. Fear may be an original tendency, but it is subject to the genetic alterations of organic growth as well as to organization by environmental conditioning. Such conditioning may determine the orientation and reference of fears but the mode of fearing undergoes change as a result of maturation. Fear is neither more nor less of an abstraction than prehension. It is not a simple entity. It waxes and alters with growth. It is shaped by intrinsic maturation as well as by experience, certainly during the period of infancy.

Consider, for example, the reactions of an infant to confinement in a small enclosed space, approximately 2 x 3 x 4 feet. In a physical sense the situation is entirely harmless. The space is ample in size; it is ventilated; it is illuminated; it is open at one end. In a personal sense, however, the space may have elements of novelty and unusualness. The infant is not accustomed to lie in such a small space which shuts him off from his accustomed environment. What are his reactions, even when he is gently introduced into this enclosed chamber? At 10 weeks he may accept the situation with complete complaisance; at 20 weeks he may betray a mild intolerance, a dissatisfaction, persistent head-turning and social seeking, which we may safely characterize as mild apprehension; at 30 weeks his intolerance to the same situation may be so vigorously expressed by crying that we describe the reaction as fear or fright. Here then are three gradations of response: first, no disquietude; second, mild disquietude, third; robust disquietude. Is not this a genetic gradation of fear behavior which is based upon maturational sequence rather than upon an historical sequence of extrinsic conditioning factors? Such factors may account for specific aspects of fear behavior, but not for the organic pattern beneath such behavior. This pattern, we would suggest, is as much the product of organic growth as the various stages in the elaboration and perfection of prehension. Incidentally it may be said that the observation of duplicate twins will tend to substantiate the existence of maturational factors in the development of emotion. Although the tendency towards developmental divergence in identical twins is probably greater in the field of personality make-up than in any other sphere of behavior, there is, during infancy, an impressive tendency toward identity of emotional behavior. Twins T and C, already referred to, showed a highly significant degree of correspondence in their manifestations of initial timidity, in their responsiveness to social games, in their reactions to the mirror image, in their gestures of avoiding and refusing, in their seeking and begging gestures, in their laughter and crying. The relatively simultaneous and progressive nature of these changes in the field of emotional behavior suggests the influence of organic maturational factors as opposed to purely extrinsic factors in the determination of behavior pattern.

THE LIMITATIONS OF CONDITIONING

The extreme versions of environmentalistic and conditioning theories suffer because they explain too much. They suggest that the individual is fabricated out of the conditioning patterns. They do not give due recognition to the inner checks which set metes and bounds to the area of conditioning and which happily prevent abnormal and grotesque consequences which the theories themselves would make too easily possible. Although it is artificial to press unduly a distinction between intrinsic and extrinsic factors, it must, after all, be granted that growth is a function of the organism rather than of the environment as such. The environment furnishes the foil and the matrix for the manifestations of development, but these manifestations come from inner compulsion and are primarily organized by inherent inner mechanics and by an intrinsic physiology of development. The very plasticity of growth requires that there be limiting and regulatory mechanisms. Growth is a process so intricate and so sensitive that there must be powerful stabilizing factors, intrinsic rather than extrinsic, which preserve the balance of the total pattern and the direction of the growth trend. Maturation is, in a sense, a name for this regulatory mechanism. Just because we do not grant complete dichotomy of internal and external factors, it is necessary to explain what keeps the almost infinite fortuities of physical and social environment from dominating the organism of the developing individual.

The organismal concept requires that the individual shall maintain an optimum or normal integrity. The phenomena of maturation suggest the stabilizing and inexpugnable factors which safeguard the basic patterns of growth. Just as the respiration of the organism depends upon the maintenance of constant hydrogen-ion concentration, so probably, on a vastly more intricate scale, the life-career of the individual is maintained by the physiological processes of maturation—processes which determine in such large measure the form and the sequence of infant-behavior pattern that the infant as an individual is reasonably secure against extreme conditioning, whether favorable or unfavorable.

The Individuality of the Infant

The foregoing emphasis of the rôle of maturation does not deprive the infant of his individuality. Maturation is by no means uniform, but displays many individual differences. The primary factors in the infant's disposition, capacities, and traits, largely determined by heredity, must be achieved through maturation. These factors represent original growth potency and are the most stable features of the infant's individuality.

Environment may play a screening or selective rôle determining which of competing potencies are to be realized and which suppressed, but the basic mechanism of realization is one of maturation. It is possible that some of the most important modifications of the developing individual are wrought by such a selective ordering. With our present slender knowledge of growth processes, this kind of modifiability is almost beyond demonstration and control. In the future more will be accomplished through an understanding of this lower, biological sphere of plasticity which is more fundamental than even the sphere of habit and conditioning.

Much of the child's individuality is undoubtedly attained through the stresses of human relations, which condition and habituate by the powerful influences of social approval. It is this extensive and pervasive interpenetration of the "environment" which falls outside of the scope of maturation as such, and which completes the patterns of organic growth. The processes of extrinsic and intrinsic organization are separated only in analytic discussion. Although the maturational factors remain primary and preponderant, the concatenation of behavior patterns, their emotional inflection and dynamic ordering lie measurably within social control. There is interdependence rather than conflict between the extrinsic and intrinsic aspect of development.

It does not follow, however, that individuality is the end product of cumulative acquired differentiation and that adults display individual differences to a greater degree than young children. There is no basis for the common opinion that infants are extremely alike. Close clinical and normative studies reveal a wealth of variation which is comparable to that of later years.

Some injustice has been done to the infant by overstressing his plasticity and conditionability. He grows in accordance with endowment as well as environment. Maturational factors preserve his native endowment. If it were not so he would become a victim of the flaccid malleability which is rather romantically ascribed to him. His own inherent equipment deserves more discriminating recognition. Only then can we do justice to the importance of individual differences which distinguish infants as well as men.

REFERENCES

These references represent different phases of investigation in the field of infancy, but are in no sense inclusive. It is impossible, within the space limitations, to list all of the authors referred to in the chapter.

1. BALDWIN, B. T., & STECHER, L. I. The psychology of the preschool child. New York: Appleton, 1925. Pp. 305.
2. BINET, A., & SIMON, T. H. The development of intelligence in children. (Trans. by E. S. Kite.) *Publ. Tr. School, Vineland, N. J.*, 1916, No. 11.
3. BLANTON, M G. The behavior of the human infant during the first thirty days of life. *Psychol. Rev.*, 1917, **24**, 456-483.
4. BÜHLER, C. Zur Psychologie des Kleinkindes. *Zsch. f. Psychol.*, 1928, **107**.
5. ————. Die ersten sozialen Verhaltungsweisen des Kindes. *Quel. u. Stud. z. Judgkd.*, 1927, No. 5. Pp. 102.
6. BÜHLER, K. Die geistige Entwicklung des Kindes. (5th ed.) Jena: Fischer, 1929. Pp. 484.
7. BURNSIDE, L. H. Coordination in the locomotion of infants. *Genet. Psychol. Monog.*, 1927, **2**, 284-372.
8. COGHILL, G. E. Correlated anatomical and physiological studies of the growth of the nervous system in Amphibia. VI. The mechanism of integration in *Amblystoma punctatum*. *J. Comp. Neur.*, 1926, **41**, No. 1. (The earlier articles of this series appeared in this journal as follows: I, **24**; II, **26**; III & IV, **37**; V, **40**.)
9. DESCOEUDRES, A. Le développement de l'enfant de deux à sept ans. Neuchatel, Paris: Delachaux & Niestle, 1921. Pp. 331.
10. GERBACH, F. M. Vocabulary studies. *Colo. Coll. Stud. Educ. & Psychol.*, 1917, No. 1. Pp. 123.
11. GESELL, A. The mental growth of the pre-school child: a psychological outline of normal development from birth to the sixth year, including a system of developmental diagnosis. New York: Macmillan, 1925. Pp. 447.
12. ————. Infancy and human growth. New York: Macmillan, 1928. Pp. 418.
13. GESELL, A., & THOMPSON, H. Learning and growth in identical infant twins: an experimental study by the method of co-twin control. *Genet. Psychol. Monog.*, 1929, **6**, No. 1. (Illustrated by a film slide of 65 photographs for visual instruction.)
14. GOODENOUGH, F. Kuhlmann-Binet tests for children of preschool age. Minneapolis: Univ. Minn. Press, 1928. Pp. 146.
15. HALVERSON, H. M. The Yale Psycho-Clinic observatory. *Amer. J. Psychol.*, 1928, **40**, 126-128.
16. JOHNSON, H. M. Children in the nursery school. New York: John Day, 1928. Pp. xx+325.
17. JONES, M. C. The development of early behavior patterns in young children. *Ped. Sem.*, 1926, **33**, 537-585.
18. KOFFKA, K. The growth of the mind. New York: Harcourt, Brace, 1924.
19. KUHLMANN, F. A handbook of mental tests: a further revision and extension of the Binet-Simon scale. Baltimore: Warwick & York, 1922. Pp. 208.
20. MARSTON, L. R. The emotions of young children. *Univ. Ia. Stud.: Stud Child Welfare*, 1925, **3**.
21. MATEER, F. Child behavior. Boston: Badger, 1918. Pp. 239.
22. MEEK, L. H., *et al.* Pre-school and parental education. *28th Yrbk. Nat. Soc. Stud. Educ.*, 1929. Pp. 831. (Contains several useful classified and abstracted bibliographies.)

23. Peiper, A. Hirntätigkeit des Säuglings. Berlin: Springer, 1928. Pp. iv+102.

24. Piaget, J. The language and thought of the child. (Trans. by M. Warden.) New York: Harcourt, Brace, 1926. Pp. 246.

25. ————. Judgment and reasoning in the child. (Trans. by M. Warden.) New York: Harcourt, Brace, 1928. Pp. 257.

26. Preyer, W. The senses and the will. (Trans. by H. W. Brown.) New York: Appleton, 1919.

27. Stern, W. Psychology of early childhood up to the sixth year of age. (Trans. from 3rd ed. by A. Barwell.) New York: Holt, 1924. Pp. 557.

28. Stutsman, R. Performance tests for children of preschool age. *Genet. Psychol. Monog.*, 1926, **1**, 1-67.

29. Thurstone, L. L. The absolute zero in intelligence measurement. *Psychol. Rev.*, 1928, **35**, 175-197.

30. Watson, J. B. Psychology from the standpoint of the behaviorist. (2nd ed.) Phildelphia: Lippincott, 1919. Pp. ix+429.

31. ————. Behaviorism. New York: People's Instit. Publ. Co, (Norton), 1925. Pp. 251.

32. Yerkes, R. M., & Foster, J. C. A point scale for measuring mental ability. (Rev. ed.) Baltimore: Warwick & York, 1923.

CHAPTER 17

THE INDIVIDUAL IN SCHOOL: I. GENERAL ABILITY

RUDOLF PINTNER
Teachers College, Columbia University

THE DEVELOPMENT OF INTELLIGENCE TESTS

Individual Differences Always Recognized. Man has always recognized the existence of individual differences in general ability among his fellow-men. This man is considered more able, more intelligent, or more brilliant than his fellows. That man is considered less able, less intelligent, or stupid. School teachers have always been aware of such differences among their pupils, even although their method of judging was poor and the judgments themselves faulty. The most frequently noticed differences among individuals were, of course, the extreme differences, and so the very weak in intellect and the very strong in intellect came in for special attention. The feeble-minded and the genius were, therefore, early set apart from the rest of mankind, and, because these two extreme groups differed so much from ordinary people, special causes to explain their existence were sought.

Tests of Simple Functions. The measurement of intelligence, however, did not arise until the coming of modern psychology. The introduction of the experimental method into psychology marks the beginning of a scientific interest in individual mental differences. From the founding of Wundt's laboratory for experimental psychology in 1879 to the construction of the first Binet Scale in 1905, we note a growing interest in the measurement of mental abilities. During the last decade of the 19th century, Cattell was writing on "Mental Tests and Measurements" and was giving tests to Columbia University students. In the Columbian Exposition in Chicago, Jastrow gave a series of tests to all people who were interested in them. Münsterberg, Bolton, Boas, Gilbert, and others were giving physical and mental tests to school children. Ebbinghaus was experimenting on his "completion test" with school children. And Binet and Henri were studying individual differences by means of the test technique. This period is one of great activity in testing. It is characterized in the main by the study of single simple tests largely psychomotor or sensory in nature. Binet, however, protests against these types of tests and contends that individual differences will be found to be more marked in the complex and higher processes.

The Work of Binet. In the next ten years we note the interest in tests continuing. Norsworthy gives group tests to discriminate between the normal and the feeble-minded child. Binet is suddenly called upon to help in the practical work of differentiating between dull and normal children in the schools of Paris, and for this purpose he brings together the results of his previous work and constructs a scale, called the 1905 Scale. This is the first scale of intelligence tests in the modern sense. The original scale was very soon revised and improved and there followed the 1908 and the 1911 Scales.

Other Influences. This very brief sketch of the early history of intelligence testing has of necessity omitted much. There were many factors at work during the 19th century leading to an interest in mental measurement. The growing interest in the unfortunate classes, notably the feeble-minded, was one such factor. The interest in heredity and eugenics aroused by the work of Galton (20-23) was another, and Galton himself conducted various simple tests. All these factors, together with the rapid growth of experimental psychology, moved along to their inevitable outcome in the measurement of intelligence. It was Binet's genius, however, that gave us the intelligence scale as we now understand it. He gave us the concept of a general intelligence that could be measured by a graded series of tests; he emphasized the importance of standardization of procedure, and he suggested the useful method of mental-age rating. For an extended treatment of the early history of intelligence testing, the reader is referred to Peterson (53).

We shall now leave the history of our subject and, without commenting upon the development of each type of test, present the tests and scales as they exist today. We shall then give the chief uses and results of the work of intelligence testing and, finally, with these results before us, we shall discuss the theory and nature of intelligence.

The Tools for the Measurement of Intelligence

Individual and Group Tests Compared. We may arbitrarily divide all tests into scales for individual testing on the one hand and group tests on the other. Individual tests are generally called scales, primarily because Binet called his first series of tests a scale. The difference between a group test and an individual test lies not only in the number of individuals measured at the same time but also in the method of testing. In the individual test the examiner is expected to put the subject at his ease, to obtain his undivided attention, and to encourage him as much as necessary. Without departing from the standard procedure of the test, the examiner must nevertheless see that as many incidental factors as possible are favorable to the subject. If the subject refuses to cooperate, gives poor cooperation, or shows nervousness, fatigue, or the like, the examiner stops the test and retests at some more favorable opportunity.

With the group test, however, such individual adjustment to the test is impossible. Only in a general way can the examiner make sure of the favorable adjustment of the group as a whole, but individuals in the group may not be in the most favorable condition for taking the test. An individual child may be fatigued or frightened or antagonistic, and such conditions may affect his score on the group test. The group test, therefore, is not as pure a measure of intelligence as the individual test. The group test contains in its score not only a measure of the intelligence of the individual but also a measure of his willingness to cooperate and put forth his best effort. Just because of this, it has proved so useful in school, because willingness to cooperate, willingness to work with the group is an important aspect of achievement in school work. There is no question here of deciding which is better, the group or the individual test. Both are good for their particular purposes.

Individual Tests. Individual tests may be roughly divided into Binet scales and performance scales. The former are direct descendants of the original Binet-Simon Scale. The latter are attempts to depart radically from the type of test found in the Binet. The Binet type is characterized by a miscellaneous group of tests given largely by means of language and depending on words and oral responses. The performance type of scale emphasizes the doing of things and uses little or no language.

The Binet scales. Here we have a miscellaneous group of tests, generally divided into age levels and generally scored plus or minus. The most commonly used scales of this type in America today are the Stanford Revision of the Binet Scale (70), the Kuhlmann Revision (46), the Herring Revision (33), and the Yerkes Point Scale (80). The last two depart from the method of age grouping. They score by points and then convert these into mental ages.

The Stanford Revision is probably the most frequently used scale, and it exhibits the chief characteristics of the group. A summary of the tests in their age grouping is as follows:

YEAR III

1. Points to parts of body—nose, eyes, mouth, hair.
2. Names familiar objects—key, penny, knife, watch, pencil.
3. Pictures—enumeration.
4. Gives sex.
5. Gives last name.
6. Repeats 6-7 syllables. (*a*) I have a little dog. (*b*) The dog runs after the cat. (*c*) In the summer the sun is hot.

YEAR IV

1. Compare lines.
2. Discrimination of forms.
3. Counts four pennies.
4. Copies square.
5. Comprehension. What must you do (*a*) when you are sleepy; (*b*) when you are cold; (*c*) when you are hungry?
6. Repeats four digits. 4-7-3-9; 2-8-5-4; 7-2-6-1.

YEAR V

1. Comparison of weights. 3 and 15 gms.
2. Colors. Red, yellow, blue, green.
3. Aesthetic comparison.
4. Definitions, use or better. Chair, horse, fork, doll, pencil, table.
5. Patience, or divided triangle.
6. Three commissions. Puts key on chair, brings box, shuts door.

YEAR VI

1. Right and left. R. hand; L. ear; R. eye.
2. Mutilated pictures.

3. Counts 13 pennies.
4. Comprehension. What's the thing to do:
    (*a*) If it is raining when you start to school?
    (*b*) If you find that your house is on fire?
    (*c*) If you are going some place and miss your car?
5. Coins. Nickel, penny, quarter, dime.
6. Repeats 16-18 syllables.
    (*a*) We are having a fine time. We found a little mouse in the trap.
    (*b*) Walter had a fine time on his vacation. He went fishing every day.
    (*c*) We will go out for a long walk. Please give me my pretty straw hat.

YEAR VII

1. Fingers. Number on R. hand; L. hand; both.
2. Pictures, description or better.
3. Repeats 5 digits. 3-1-7-5-9; 4-2-8-3-5; 9-8-1-7-6.
4. Ties bow knot. Model shown. Single bow half credit.
5. Gives differences. (*a*) fly and butterfly, (*b*) stone and egg, (*c*) wood and glass.
6. Copies diamond.

YEAR VIII

1. Ball and field. (Inferior plan or better.)
2. Counts backward 20 to zero.
3. Comprehension. What's the thing for you to do:
    (*a*) when you have broken something which belongs to someone else?
    (*b*) when you are on your way to school and notice that you are in danger of being tardy?
    (*c*) if a playmate hits you without meaning to do it?
4. Similarities. (*a*) wood and coal; (*b*) apple and peach; (*c*) iron and silver; (*d*) ship and automobile.
5. Definition superior to use. (*a*) balloon; (*b*) tiger; (*c*) football; (*d*) soldier.
6. Vocabulary, 20 words.

YEAR IX

1. Date: week, month, day of month, year.
2. Weights.
3. Makes change. 10-4; 15-12; 25-4.
4. Repeats 4 digits backwards. 6-5-2-8; 4-9-3-7; 8-6-2-9.
5. Three words in a sentence. (*a*) Boy, river, ball; (*b*) work, money, men; (*c*) desert, rivers, lakes.
6. Rhymes. Three rhymes for each word. (*a*) day; (*b*) mill; (*c*) spring.

YEAR X

1. Vocabulary, 30 words.
2. Absurdities.
    (*a*) A man said, "I know a road from my house to the city which is down hill all the way to the city and down hill all the way back home."
    (*b*) An engineer said that the more cars he had on his train the faster he could go.
    (*c*) Yesterday the police found the body of a girl cut into 18 pieces. They believe that she killed herself.
    (*d*) There was a railroad accident yesterday, but it was not very serious. Only 48 people were killed.
    (*e*) A bicycle rider, being thrown from his bicycle in an accident, struck his head against a stone and was instantly killed. They picked him up and carried him to the hospital, and they do not think he will get well again.
3. Designs.

4. Reading and report.
   New York. | September 5th. | A fire | last night | burned | three houses | near the center | of the city. | It took some time | to put it out. | The loss | was fifty thousand dollars, | and seventeen families | lost their homes. | In saving | a girl | who was asleep | in bed, | a fireman | was burned | on the hands.
5. Comprehension.
   - (*a*) What ought you to say when someone asks your opinion about a person you don't know very well?
   - (*b*) What ought you to do before undertaking (beginning) something very important?
   - (*c*) Why should we judge a person more by his actions than by his words?
6. Sixty words in three minutes.

YEAR XII

1. Vocabulary, 40 words.
2. Abstract words. (*a*) pity; (*b*) revenge; (*c*) charity; (*d*) envy; (*e*) justice.
3. Ball and field. (Superior plan.)
4. Dissected sentences.
   - (*a*) for the started an we country early at hour.
   - (*b*) to asked paper my teacher correct I my.
   - (*c*) a defends dog good his bravely master.
5. Fables.
6. Repeats 5 digits backwards. 3-1-8-7-9; 9-6-4-8-2; 5-2-9-6-1.
7. Pictures, interpretation.
8. Gives similarities. (*a*) snake, cow, sparrow. (*b*) book, teacher, newspaper. (*c*) wool, cotton, leather. (*d*) knife-blade, penny, piece of wire. (*e*) rose, potato, tree.

YEAR XIV

1. Vocabulary, 50 words.
2. Induction test. (Gets rule by 6th folding.)
3. President and king. Difference between.
4. Problems of fact.
   - (*a*) A man who was walking in the woods near a city stopped suddenly, very much frightened, and then ran to the nearest policeman, saying that he had just seen hanging from the limb of a tree a ————————. A what?
   - (*b*) My neighbor has been having queer visitors. First a doctor came to his house, then a lawyer, then a minister. What do you think happened there?
   - (*c*) An Indian who had come to town for the first time in his life saw a white man riding along the street. As the white man rode by, the Indian said, "The white man is lazy; he walks sitting down." What was the white man riding on that caused the Indian to say he was sitting down?
5. Arithmetical reasoning.
   - (*a*) If a man's salary is $20 a week and he spends $14 a week, how long will it take to save $300?
   - (*b*) If 2 pencils cost 5c, how many pencils can you buy for 50c?
   - (*c*) At 15c a yard, how much will 7 feet of cloth cost?
6. Clock. Reverse hands; (*a*) 6:22 (*b*) 8:10 (*c*) 2:46.

YEAR XVI

1. Vocabulary, 65 words.
2. Fables.
3. Difference between abstract words. (*a*) laziness and idleness; (*b*) evolution and revolution; (*c*) poverty and misery; (*d*) character and reputation.
4. Problem of the enclosed boxes. One large box containing (*a*) 2 smaller,

1 inside of each; (*b*) 2 smaller, 2 inside of each; (*c*) 3 smaller, 3 inside of each; (*d*) 4 smaller, 4 inside of each.

5. Repeats 6 digits backwards. 4-7-1-9-5-2; 5-8-3-2-9-4; 7-5-2-6-3-8.
6. Code. Writes, "Come quickly."

YEAR XVIII

1. Vocabulary, 75 words.
2. Paper-cutting test. Draws and locates holes.
3. Repeats 8 digits forwards. 7-2-5-3-4-8-9-6; 4-9-8-5-3-7-6-2; 8-3-7-9-5-4-8-2.
4. Repeats thought of passage heard.
   - (*a*) "Tests such as we are now making are of value both for the advancement of science and for the information of the person who is tested. It is important for science to learn how people differ and on what factors these differences depend. If we can separate the influence of heredity from the influence of environment we may be able to apply our knowledge so as to guide human development. We may thus in some cases correct defects and develop abilities which we might otherwise neglect."
   - (*b*) "Many opinions have been given on the value of life. Some call it good, others call it bad. It would be nearer correct to say that it is mediocre, for on the one hand our happiness is never as great as we should like, and on the other hand our misfortunes are never as great as our enemies would wish for us. It is this mediocrity of life which prevents it from being radically unjust."
5. Repeats 7 digits backwards. 4-1-6-2-5-9-3; 3-8-2-6-4-7-5; 9-4-5-2-8-3-7.
6. Ingenuity test.
   - (*a*) A mother sent her boy to the river to get 7 pints of water. She gave him a 3-pint vessel and a 5-pint vessel. Show me how the boy can measure out exactly 7 pints without guessing at the amount. Begin by filling the 5-pint vessel.
   - (*b*) Same, except 5 and 7 given to get 8. (Begin with 5).
   - (*c*) Same, except 4 and 9 given to get 7. (Begin with 4).

The fables used in Ages XII and XVI are entitled (*a*) Hercules and the Wagoner; (*b*) The Milkmaid and Her Plans; (*c*) The Fox and the Crow; (*d*) The Farmer and the Stork; (*e*) The Miller, His Son, and the Donkey. A score of two is given for a generalized interpretation, and a score of one for an interpretation that does not go beyond the characters in the fable.

The words used in the Vocabulary Test consist of two series of fifty words each. The idea of the test is to find out whether the child knows the words and not whether the child can define them accurately. A word is considered correct if the subject shows he knows any meaning of the word, however poorly he may define it. The arrangement of the words in the list is from very easy to very hard.

The performance scales. As a type of the miscellaneous performance scale we shall describe the Pintner-Paterson Performance Scale (57). It consists of 15 tests as follows:

1) Mare and Foal Board. This is a picture board of a mare and foal with a number of cut-outs which the subject has to put in the correct places. It is very simple, resembling a child's game, and serves as a very good introduction for children. Time and number of errors are recorded.

2) Seguin Formboard. Ten blocks representing common geometrical forms are to be placed in their appropriate places. The time of the shortest of three trials is recorded.

3) Five-Figure Board. Five geometrical figures each divided into two or three pieces are to be placed in their appropriate places. Time and number of errors are recorded.

4) Two-Figure Board. Nine pieces to be placed in two spaces. Time and number of moves are recorded.

5) Casuist Board. A more difficult board, consisting of four spaces into which have to be fitted twelve blocks. Time and number of errors are recorded.

6) Triangle Test. Four triangular pieces to be fitted into the board. Time and errors are recorded.

7) Diagonal Test. Five variously shaped pieces are to be fitted into a rectangular frame. Time and errors are recorded.

8) Healy Puzzle A. Five rectangular pieces are to be fitted into a rectangular frame. Time and moves are recorded.

9) Manikin Test. Subject has to put together legs, arms, head, and body to form a man. There is no board into which the pieces fit. Quality of performance is scored.

10) Feature Profile Test. In the same manner as in the previous test, subject has to put together pieces to form a head. Time is recorded.

11) Ship Test. This consists of the picture of a ship cut into ten pieces of the same size and shape which are to be fitted together properly in a rectangular frame. Quality of performance scored.

12) Picture Completion Test. Subject is required to select the appropriate block out of many possible blocks to complete the picture. Quality of performance scored.

13) Substitution Test. A sheet of paper with rows of geometrical figures upon which the subject has to write the proper digit following the key at the top of the page. Time and errors compounded into a score.

14) Adaptation Board. This is a simple test for measuring the ability of the subject to keep his attention upon a moving board. Number of correct moves is recorded.

15) Cube Test. Four cubes are tapped in a certain order and the subject is required to watch and then imitate the movement. Number of combinations correctly imitated is recorded.

The special feature of this scale is that it can be given without the use of any language and no language is required for any of the tests. It is, therefore, indispensable in testing foreign subjects, deaf children, and cases of limited verbal development. It should be used with all children to test those other aspects of intelligence which are not taken care of by the Binet type of scale. A correlation of Binet and performance M. A.'s of .83 has been reported by Johnson (40). Figure 1 shows the ten tests of the Pintner-Paterson Short Performance Scale.

Other performance scales are:

1) The Army Performance Scale (81). This is a series of performance tests of the same type as the Pintner-Paterson tests. It was used in the testing of foreign and illiterate recruits in the army in 1917-18.

2) The Porteus Maze Scale (58). This consists of a graduated series of mazes, one maze for each age from age 3 to age 12 inclusive and also one for age 14. Unlike the preceding performance scales, this scale limits itself to one type of reaction, namely, maze tracing.

3) The Dearborn Form Boards (11). This is a series of formboards of different degrees of difficulty.

4) The Worcester Form Board Series (62). This is another series of formboards graduated in difficulty from very easy to hard.

5) The Kohs Block Design Scale (45). This uses a set of color cubes. The subject constructs the design placed before him. There are 17 designs, increasing in difficulty from easy to hard.

6) The Arthur Point Performance Scale (1). This makes use of seven of the Pintner-Paterson tests and the Kohs Block Design Scale.

7) The Goodenough Drawing of a Man (26). Perhaps this should

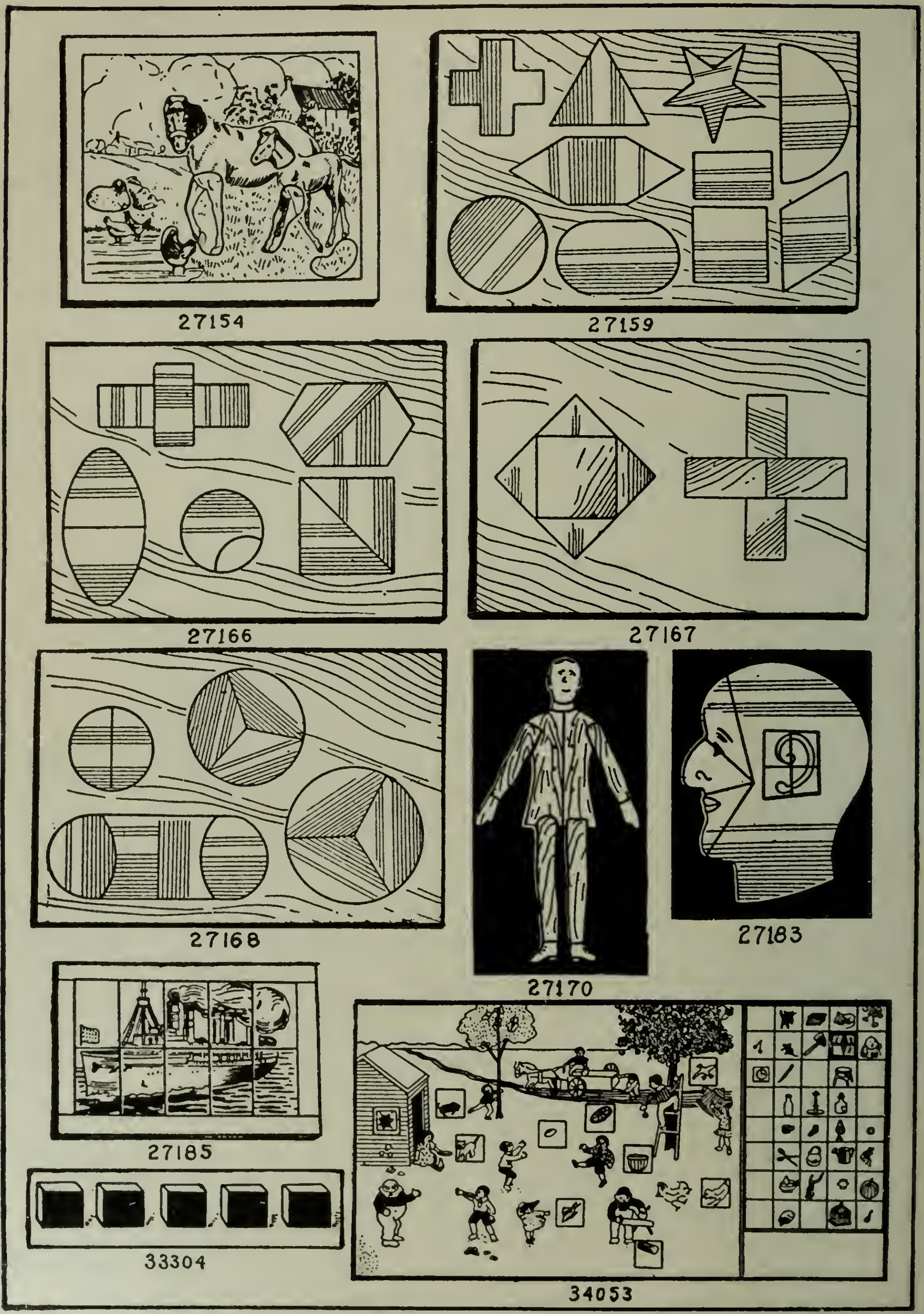


FIGURE 1
The Pintner-Paterson Short Performance Scale
(Courtesy of C. H. Stoelting Company, Chicago)

not be included among performance tests proper. It is, nevertheless, of the performance type. The subject is told to draw a man. There is a standard method of scoring the drawing.

In addition to these regular scales there are numerous performance tests and partially standardized scales (5). There are also several special scales for the testing of very young children, below school age. There are tests for very young children in the Kuhlmann Revision of the Binet, but recently attempts have been made to construct special scales, more of the performance type, for such cases. The most commonly used is the Merrill-Palmer or Stutsman (67) series of tests. Again important for the intelligence measurement of young children is the Gesell (25) series of developmental levels from birth to age five.

We see, therefore, that the psychologist possesses a great number of scales for the individual measurement of intelligence, ranging all the way from the highly linguistic type of scale of the Binet variety to the non-language performance type, with all kinds of intermediate mixtures. And again, these scales measure levels of intelligence from very low to fairly high. The greater number of these scales are suitable for the intermediate levels of mentality corresponding to mental ages of six to twelve. Lower levels of mentality are taken care of by the recent scales for young children and infants. For mental levels above age 12 the Binet type of scale must at present be used, and for very superior intellect this soon becomes inadequate. For high-school children of superior intellect, for college students and the like, we have no adequate individual tests of intelligence. All these cases can, however, be effectively measured by group tests.

Group Tests. Group tests may profitably be discussed from the point of view (1) of content of material in the test; and (2) of the level or difficulty of the test, or, in other words, the range of mental ability that can be measured by any given test.

Content. The material used in group tests consists of verbal material, numbers, diagrams, pictures. Most tests use verbal material almost entirely, and these are called verbal tests. Some tests use pictures or diagrams entirely, but the subject is given oral instructions as to what to do, and these are called non-verbal tests, because the content is non-verbal. Some tests use pictures and diagrams, but the subject is told what to do by means of samples and by means of pantomime. No words are used in the test or in the directions for doing the test. These tests are called non-language.

Verbal tests are by far the most common. They are made up of different sorts of verbal material. Some of the common types are:

(A) *Opposites.* This is one of the oldest and most useful tests. The subject is called upon to respond by writing down or indicating the opposite of a given word, or by deciding whether two words denote opposite or similar ideas.

Underline the word in parenthesis which is the opposite of the first word:
accept. . . . .(receive, percept, deny, reject, spend).
constant. . . . (always, fickle, stationary, seldom, movable).

Underline "opposite" if the two words mean the opposite, and "same" if they mean the same:

furtive .... sly same .... opposite
any .... none same .... opposite
asunder .... apart same .... opposite
deplete .... exhaust same .... opposite
superfluous .... essential same .... opposite

(B) *Analogies.* This test has proved one of the most valuable. The analogy between two words is given and the subject has to decide as to a similar analogy with reference to another pair.

Underline the best of the four words in parenthesis:

cellar ....attic bottom..(well, tub, top, house)
man......arm tree....(shrub, limb, flower, bark)
imitate....copy invent..(study, Edison, machine, originate)
physics..motion (?)..blood..(temperature, body, physiology, geography)

(C) *Best Reasons.* This test apears in many forms. It is often called a test of common sense, or comprehension. The subject indicates in some form or other the best answer to a question.

Check the best reason:

The cause of echoes is
(*a*) the reflection of sound waves
(*b*) the presence of electricity in the air
(*c*) the presence of moisture in the air
Gold is more costly than lead, because
(*a*) it is of finer appearance
(*b*) it is more scarce
(*c*) it is used more for jewelry.
(*d*) it is yellow

(D) *Disarranged Sentences.* This seems to have had its origin in the Binet tests. The words in a sentence are disarranged, and the subject has to arrange them properly.

Underline "true" or "false" according to the meaning of the disarranged sentence:

will live bird no forever true..false
always sleeplessness clear causes a conscience true..false

Cross out the superfluous word in the disarranged sentence:

watch summer the man stole is jail who the in
bushes trees hay roots have and their the ground in

(E) *Proverbs.* The subject has to match proverbs of similar meaning, or decide whether they are the same or different in meaning, or match them with statements which explain the meaning.

Proverbs.

1. The burnt child dreads the fire.
2. Rome was not built in a day.
3. There is no smoke without fire.

etc.

Mark the statements which explain these proverbs.

Time is required to produce anything of value.
Failure follows frequent change of plan.
Unhappy experiences teach us to be careful.
Those in disgrace always want to disgrace others.
There is no result without a cause.

(F) *Number Completion.* This calls for discovering the rule or method in the arrangement of a series of numbers and indicating this in some way.

Write down the two numbers that should come next:

3 4 6 9 13 18 — —
21 18 16 15 12 10 — —

Fill in the missing numbers:

26 22 — 14 10 — 2
72 — — — — 37 — — — — 2

Cross out the number that does not belong in the series:
2 4 8 10 16 32
72 36 18 9 6

(G) *Directions.* This is one of the earliest types of group test. The subject is asked to do just what he is told:

Cross out the "g" in tiger.
Put a dot below this line—.
If Decoration Day comes in winter write the word "No." If not, write the word "Yes."
Write the letter which follows the letter which comes next after C in the alphabet.

(H) *Sentence Completion.* Words omitted in a sentence or passage have to be filled in.

Write one word on each blank:
The boy ——— two dollars to the Red Cross.
Those things ——— ——— no fear ——— sometimes ——— harmful.

Underline the right word of the three corresponding to the number in the blanks

Once upon a (1) there was a young (2) who was very (3) . He went from (4) to (4) trying to find (5) .

1.	time	place	cat
2.	bird	man	woman
3.	rich	strong	poor
4.	place	there	snow
5.	him	gold	work

(I) *Information.* The subject is required to show his general information. Usually the items are spread over a wide field. In so far as the material is based on school knowledge, it is an attempt to get at the intelligence of an individual by measuring what has been learned and retained. This may at times give us wrong measures, particularly if we compare children of very different environments.

Underline the correct word:
Euchre is played with dice, rackets, cards, pins.
The Delco System is used in plumbing, filing, ignition, cataloguing.
Coral is found in trees, reefs, mollusks, mines.
John Wesley was famous in literature, science, war, religion.

Non-verbal tests use pictures or diagrams. As an example we may take the Detroit First-Grade Intelligence Test. Test 1 shows pictures of a spoon, scissors, hammer, knife, and the subject is told to mark two things we cut with. Test 4 shows pictures with things that are wrong and the child is told to "mark the wrong thing in the sink picture; in the basket picture"; and so on. Test 6 shows pictures of a hat, a shoe, and a glove, and the instructions are, "Mark the thing you put on your foot." A child, therefore, does not need to be able to read or recognize verbal symbols of any kind, but he must understand the English language and know the meanings of a lot of words to be able to do the test at all.

The non-language type of test is, of course, non-verbal in content and in addition does not presuppose a knowledge of any language. As an example we may take the Pintner Non-Language Test. Test 1 has twelve lines of four dots each. The examiner draws a line of dots on the blackboard and moves his pointer from dot to dot in various combinations. These movements of the pointer are then reproduced in pencil by the pupil on his test sheet. Test 4 shows incomplete pictures which the child has to complete. Test 4 is shown in Figure 2.

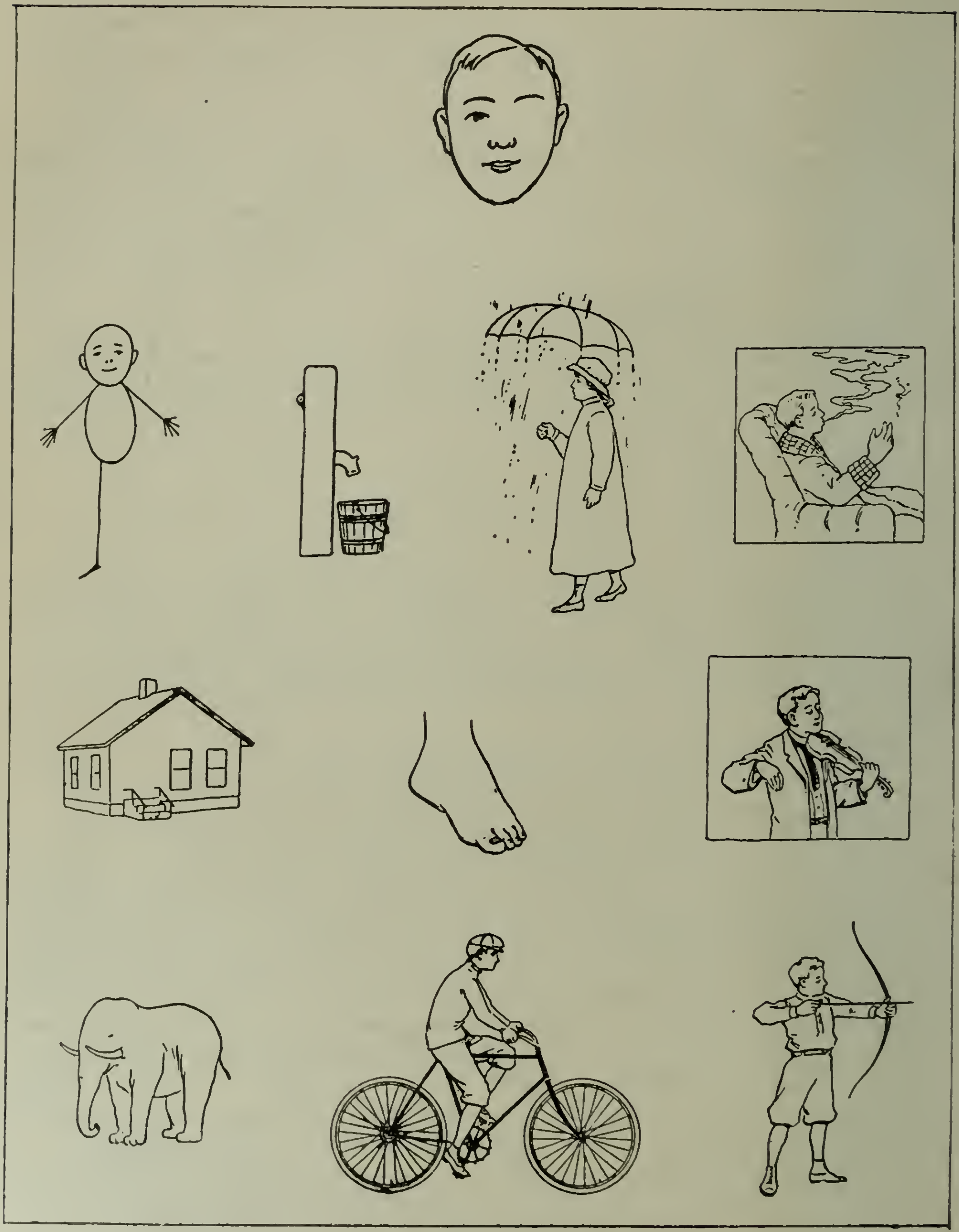


FIGURE 2
THE COMPLETION TEST FROM THE PINTNER NON-LANGUAGE MENTAL TEST

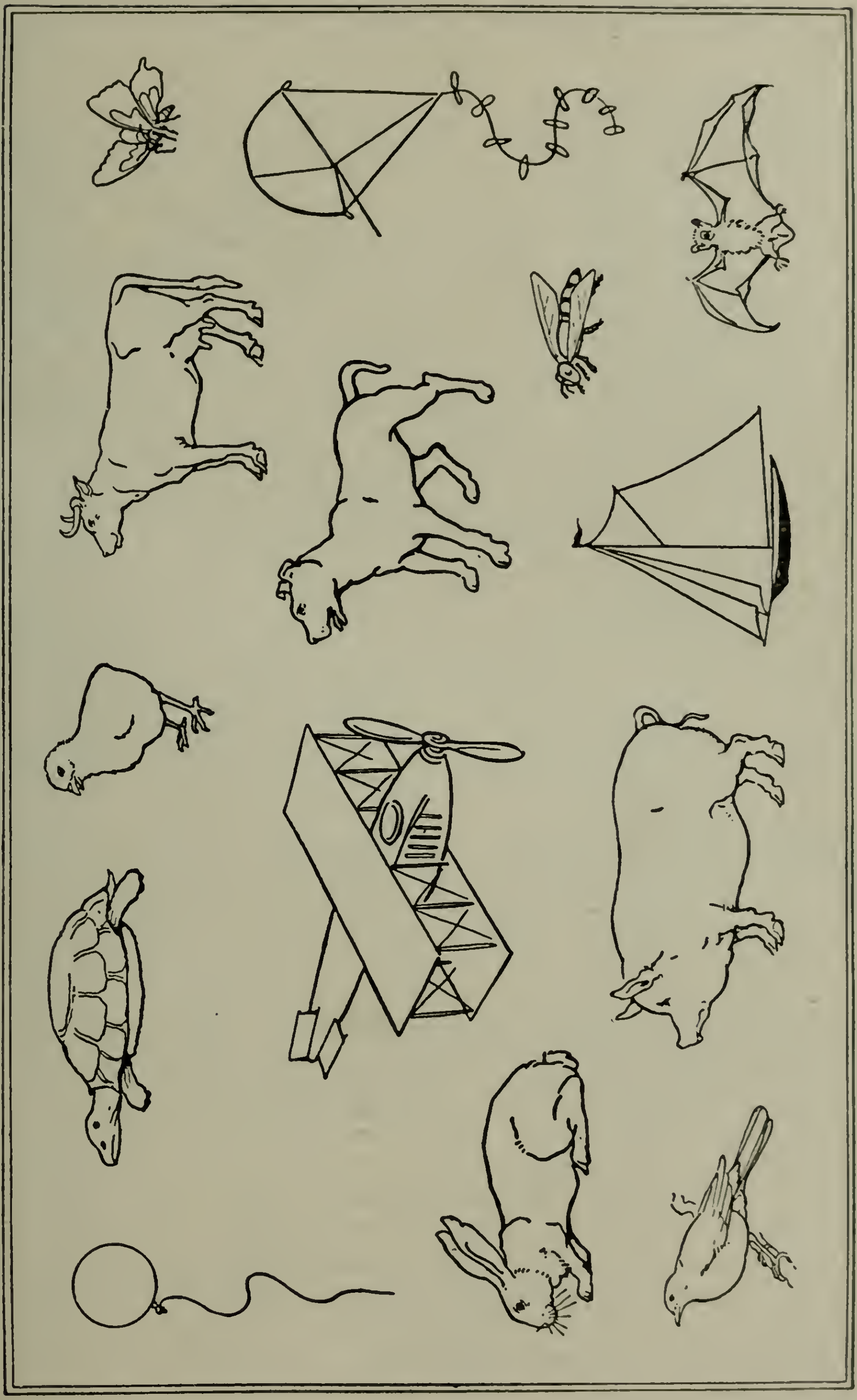


FIGURE 3

A Page from the Pintner-Cunningham Primary Mental Test

(Courtesy of World Book Company, Yonkers)

2) *Level.* Any given test will measure only a certain range of mentality. Material appropriate for low levels of mental ability will not measure high levels and vice versa. Hence we have different tests suitable for different levels.

The kindergarten or first-grade level may be illustrated by the Pintner Cunningham Primary Mental Test. This is non-verbal as to content. There are no words nor numbers. Figure 3 shows page 3 of the booklet. The children are told to put a mark on all those things which go up in the air. Test 2 shows series of three pictures and they are told to put a mark on the prettiest. Test 4 shows a doll with various dresses, hats, shoes, etc., and the child is told to put a mark on the right dress, shoe, etc.

The National Intelligence Test is suitable for a higher level of mentality, such as we find in Grades 4 to 8 in the elementary school. The sub-tests for Scale A are (1) Arithmetic Reasoning; (2) Completing Sentences, such as *The dog——black;* (3) Common Attributes, e.g., underline the two words which tell what a mouse always has—*back, cat, eyes, cheese, trap;* (4) Similarities and Differences, e.g., *new* and *old, fall* and *drop, expect* and *anticipate;* (5) Symbol-Digit, in which the subject copies the digits for the appropriate symbols as shown in a key at the top of the test page.

The high-school level of mentality may be illustrated by the Terman Group Intelligence Test. It is entirely verbal in content and consists of ten tests: (1) Information; (2) Best Answer; (3) Word Meaning; (4) Logical Selection; (5) Arithmetic Reasoning; (6) Sentence Meaning; (7) Analogies; (8) Mixed Sentences; (9) Classification; (10) Number Series.

At the college or superior adult level we may take the Thorndike Intelligence Examination as an example. It consists of a battery of four tests. The first two contain much the same type of material already described, only more difficult. The last two contain difficult reading passages.

We see, therefore, that there are appropriate group tests for all levels of mental ability from the kindergarten up to college. The content varies at different levels, being non-verbal at the lower levels and strictly verbal at the upper. There are no adequate non-language tests at the lower or at the upper levels. There is only one test that measures intelligence by means of the same kind of content all the way from a low to a high level of intelligence. This is the Thorndike CAVD Test. It is given as an individual test at the lowest levels and as a group or individual test at intermediate and higher levels. The letters CAVD refer to the content of the test, namely, Completion, Arithmetic, Vocabulary, and Directions. It is calibrated better than any other test. It proceeds by steps, equal in difficulty, from a low to a high level.

List of Group Tests. Below is a list of commonly used group intelligence tests arranged roughly according to level. It is by no means complete, but it shows those which have been published and used frequently.

GROUP INTELLIGENCE TESTS

Arranged according to grades for which they are more or less suitable

Kindergarten	Pintner-Cunningham Primary Test	(World Book)
	Kingsbury Primary Test	(Pub. School Publ. Co.)
	Rhode Island Intelligence Test	(Pub. School Publ. Co.)
Grades 1—3	Pressey Primary	(Pub. School Publ. Co.)
	Dearborn Tests I	(Lippincott)
	Haggerty Delta I	(World Book)
	Detroit Primary	(World Book)
	Otis Primary	(World Book)
	Pintner-Cunningham	(World Book)
Grades 4—8	Haggerty Delta II	(World Book)
	Dearborn Tests II	(Lippincott)
	McCall Multi-Mental Scale	(Teach. Coll. Bur. Publs.)
	National Intelligence Scale	(World Book)
	Otis Group	(World Book)
	Chicago Group Intelligence	(Univ. Chicago)
	Pintner Non-Language	(College Book Co., Columbus, O.)
	Illinois Examination	(Pub. School Publ. Co.)
	Otis Self-Administering (Intermediate)	(World Book)
	Pintner Rapid Survey Test	(Teach. Coll. Bur. Publs.)
High school	Terman Group	(World Book)
	Miller Mental Ability	(World Book)
	Otis Self-Administering Examination (Higher)	(World Book)
	Thurstone Intelligence	(Stoelting)
	Army Alpha	(Stoelting)
College	Thorndike College Entrance	(Teach. Coll. Bur. Publs.)
	Brown University Test	(Lippincott)
	Revised Army Alpha	(Psychol. Corp.)
	Roback Mentality Tests	(Miller, Medford, Mass.)
	Army Alpha	(Stoelting)
	Ohio State University Test	(Professor Toops, Ohio State Univ.)

The following tests have forms which are applicable to all grades from kindergarten through university.

CAVD	(Instit. Educ. Res. Teach. Coll.)
Kuhlmann-Anderson Intelligence Tests	(Test Bur., Minneapolis, Minn.)
Myers Mental Measure	(Newson, New York)
Trabue Mentimeter	(Doubleday, Page, New York)

Evaluation of Scores. After giving an intelligence test, the result one gets is generally a score of so many points. This, of course, is meaningless without further interpretation. The most common method of evaluation is to turn the score into a mental age or M.A. by means of a table of norms. This means a comparison of the child examined with the group of children upon which the test was standardized. If the score approximates the average score of the 10-year-old standardization group, then the score is said to be equivalent to a mental age of 10, and so on for other mental ages. This mental-age scheme of evaluation is simple and readily comprehended. It assumes, however, that the groups used for the standardization of different tests are very similar. If they are not, the same child might obtain several different mental ages on different tests, not because his real intelligence fluctuates from test to test, but simply because the standardization groups of the tests in question were not similar.

Some of the revisions of the Binet arrive at the M.A. directly without going through the process of computing a score in points and then changing this score into a mental age.

The M.A. is a statement of the level of intelligence attained by a given subject at a given time. If we know that an individual has an M.A. of 6, we know that his level of intelligence is low; we do not know, however, whether his intelligence is inferior or superior. This we can arrive at only by a comparison of his level with the level he should normally have attained. This is done by comparing his mental age with his chronological age, and the ratio of M.A. to C.A. gives the intelligence quotient. The intelligence quotient or I.Q. is thus a measure of brightness. It tells whether the individual possesses superior, average, or inferior intelligence for his age. Furthermore, to the extent that the I.Q. is constant from year to year, the I.Q. is an indication of the rate of growth of intelligence. Children with high I.Q.'s have grown rapidly intellectually and may continue to do so; those with low I.Q.'s have grown slowly.

The M.A.-I.Q. method of evaluating intelligence scores is very easy to understand and is in common use. It has, however, distinct limitations. It can really be used only up to the point where intelligence stops growing. This point, as we shall see later, has not been actually determined. At any rate it can readily be seen that if we used the I.Q. method above this point, we should get steadily decreasing I.Q.'s as the individual grew older. For example, if the average adult M.A. is 18, then an average adult of that M.A. could not obtain a higher M.A. There could be no higher M.A., if M.A. is the average performance of individuals of each chronological age group. The average scores of the chronological age groups of 18, 19, 20, and so on would all be the same. If we consider such an average score as equivalent to an M.A. of 18, then for our calculation of the I.Q. we must use 18 for all individuals whose age is 18 or above. If not, if we use their real chronological ages, then the average individual of 20 or 25 or 30 would obtain an I.Q. decidedly below 100. Furthermore, if an individual of high intelligence obtains a score above this point of average adult mentality, we cannot assign him a real M.A. because there is no such M.A. in the sense that we have been using it, i.e., the average performance of a given chronological age group. If increase in scores by age groups stops at 18, or any other given age, then the average score for all other age groups above this is the same, and we cannot have higher M.A.'s. In order, therefore, to evaluate high scores above any given point at which a scale stops measuring average individuals, fictitious M.A.'s are frequently assigned. It should be distinctly understood, however, that these M.A.'s are merely fictions in order to allow the calculation of I.Q.'s for high scores. They are not based upon the average score for a chronological age group, as is the case with the lower M.A.'s. Hence the M.A.-I.Q. method breaks down in the upper ages of adolescence. It is a useful scheme for younger children; it is of doubtful value in the high school and of no value above that level.

A much sounder method of evaluating intelligence scores is to interpret the score in terms of the variability of the group. This is frequently called the standard-measure method.

$$Z = \frac{X - M}{\sigma}$$

where Z is the standard measure, X the individual score, M the mean of the group, and σ the standard deviation. The T-score method of McCall (49) is a special case of this standard-measure method. The T unit is one-tenth of the standard deviation of an unselected 12-year-old group used for standardization purposes. Tables of T scores have been arranged for many tests.

The Uses and Results of Intelligence Testing

Classification. The most common use of intelligence tests is as an aid to classification of children in school. Children differ greatly in the amount or level of intellect and they also differ greatly in rate of growth or I.Q. Other things being equal, it is of advantage to group children of the same level and rate together for instructional purposes. Homogeneous groups can be taught better than heterogeneous groups.

Mental Age to Determine Grade. The grade at which a child can do his most effective work will be in part determined by his mental age or level of intelligence. First-grade work is more or less adjusted to six-year-old mentality; second-grade, to seven-year-old mentality; third-grade, to eight-year-old mentality; and so on. So far as possible, therefore, children should be placed in grades appropriate to their mental ages. This does not mean that no other factors should be taken into account. Of course they should. Health is important. Social development is important, particularly when we come to the high-school level. Educational attainment must be considered. For example, if we find a seven-year-old child with an M.A. of 10 years, 6 months, and an I.Q. of 150 in the second grade, his mental level would indicate his ability to do the school work of the fifth grade. No psychologist would recommend his immediate promotion to the fifth grade, but if his health were good and the school had no special classes for bright children, then immediate promotion to the third grade would be indicated with acceleration to the fourth grade in the near future. Too much attention should not be paid to present educational achievement, because many bright children are satisfied to do mediocre work in the grades in which they find themselves, and an extra promotion frequently acts as a potent stimulus to do better work.

Proper classification is very important in the kindergarten and first grade. The first year or two of school life frequently determine the amount of effort a child will put into his school work. A very bright child with mental ability well beyond that required for first-grade work, if left in the first grade, will not need to exert any mental effort to accomplish the work. He receives commendation from the teacher for doing easy tasks and so learns to be satisfied with that type of performance. And so he may continue during his scholastic career, satisfied with work far below the level of his ability.

In this general way, therefore, the mental level will help to determine the proper grade. Bright children can be accelerated. Dull children can be prevented from being pushed ahead too fast, so that their school work becomes a burden to them and their only wish is to leave school as soon as possible. The greatest satisfaction is achieved when children are working at their true level of ability.

The I.Q. for Sectioning. In large schools, where there are several sections of each grade, a further refinement in classification can be made. Supposing we have placed all our children in the most appropriate grades according to their level of ability and a due consideration of any other important factors, then we will find that the children in any one grade will vary greatly in I.Q. There will probably be a scattering of I.Q.'s all the way from about 80 to about 150. The actual range of I.Q.'s in any particular case will be determined by the neighborhood in which the school is located. The frequency distribution of the I.Q.'s of a 5A grade tested on the National Intelligence Test is given in Table 1.

TABLE 1

I.Q.	Frequency
130-139	2
120-129	2
110-119	14
100-109	16
90-99	22
80-89	18
70-79	11
60-69	7
50-59	1
Total	93

Now, even if all these children have the same M.A. or mental level at the beginning of the year, and therefore are capable of doing the same work, yet the various I.Q.'s indicate various rates of growth, various rates of ability to progress with their school work. To the extent that the I.Q. is constant, the child with the I.Q. of 50 will grow only half a year in mentality during the year; the child with an I.Q. of 100 will grow a whole year; and the child with the highest I.Q. will grow a year and a third. If, therefore, we pay no attention to their I.Q.'s in classification, we shall have increasing differences in mental level as the year progresses. The natural thing to do, therefore, is to divide them into sections according to their I.Q.'s, so that the classes will be as homogeneous as possible from the point of view of rate of growth of mental ability. In the example given above, three rooms and three teachers were available. The

30 highest I.Q.'s were put in one section, the next 35 in another section, and the lowest 28 in another section. Just where the lines should be drawn for the different sections will always depend upon the number of rooms, the size of rooms, the number of teachers, and so on. In general it is better to make the highest and lowest sections somewhat smaller than the intermediate sections, if possible, because the differences in I.Q.'s will be greater in the former than in the latter.

Special Classes for Extreme Cases. The general scheme of classification outlined above does not take adequate care of extreme cases. Children of extremely high intellect would be so far advanced beyond their mates of similar chronological age, and children of extremely low intellect would be so far retarded, that difficulties of adjustment would undoubtedly take place. A seven-year-old boy with an M.A. of 11-7 and an I.Q. of 165 in the sixth grade, or a thirteen-year-old boy with an M.A. of 7-7 and an I.Q. of 58 in the second grade would be examples of socially maladjusted children. For such extreme cases special classes are desirable.

For the dull and feeble-minded such classes have been long in existence and they are a recognized part of most school systems. Generally the children in these classes have I.Q.'s from about 50 to about 70 or 80. Such children, making one-half to two-thirds of normal progress each year, need special attention during their whole school career. From the very beginning they should be handled differently, so that they do not get a pronounced dislike of school. They must be given more concrete things to do and they should be prepared for some suitable vocation by the time they are ready to leave school. Assignment of children to such a special class should not be made on the basis of a group intelligence test. A thorough individual examination by a competent examiner is necessary. The children should be frequently retested.

For children of extremely high intellect, there are as yet very few special classes. There is a tendency to think that they can take care of themselves without special help and that they will come to the top anyway, regardless of their schooling. On the other hand, the brilliant achievements of such children when put into special classes show what they are capable of doing when given adequate opportunities. It is a waste of good intellect to allow them to mark time in the ordinary grades. A certain amount of acceleration will help, but acceleration alone soon raises acute problems if it is carried too far, problems of the adjustment of the young bright child with the older normal child. Special classes for the very bright seem the best solution for these problems.

Terman's (72) extensive study of children with I.Q.'s above 135 shows that such children are in general taller, heavier, and healthier than the average child. They are just as much interested in games, but they read much more extensively. They are not freaks or prodigies in any sense of those terms. They are rarely conscious of their superior intellect and therefore are seldom vain or conceited. If they are, such undesirable characteristics are speedily cured by a special class for bright children, where they have to compete with others of equal or greater ability. For the identification of children of high intelligence for the purpose of forming special classes, the use of group or individual intelligence tests is very necessary. School grades and teachers' judgments are not very reliable in discovering the child of high intellect.

Guidance. Since, as we shall see later, the I.Q. is relatively stable, it is in a sense a prognostic measure. A child with a high I.Q. in the elementary school is likely to have a high I.Q. in high school and also in college. The chances are, therefore, that he will be able to do those things which demand high intellect. In this narrow sense, therefore, intelligence tests are of help in educational and vocational guidance. But the intelligence rating is only one of many factors that must be considered in advising a child as to his future career, and it is more useful in the narrow realm of educational guidance, than in the broader field of vocational adjustment. Most of our abstract intelligence tests are closely correlated with school attainment. They are measuring the sort of ability necessary for the mastery of the academic subjects. Hence we may make certain general predictions. A child with an I.Q. of 80 or less is not likely to complete the eighth grade successfully. By the time he is 14 or 16 his mental age will be about 11 or 12. He should therefore be trained for the more concrete tasks of the world. A child with an I.Q. below 90 will probably derive little profit from the ordinary high-school course. It needs an I.Q. of about 110 and above to master the abstractions and make the generalizations required in algebra (74). Children in high school with I.Q.'s between 90 and 110 will have difficulty unless their studies are carefully chosen. When it comes to advice with reference to future college work, it would seem that an I.Q. of at least 110 is required for the usual academic work in a liberal arts college.

The College Student. Intelligence tests have been given to great numbers of college students, and in many colleges and universities they are in regular use. MacPhail (48) lists the following 13 uses that have been made of test results:

1) To predict the probable quality of academic work that each student will do.

2) To select and refuse entrance to those applicants who are not sufficiently gifted in intelligence to do college work.

3) To serve as a regular means of entrance supplementary to or taking the place of certain customary entrance examinations.

4) To determine admission of special cases who cannot fulfill the usual entrance requirements.

5) To classify students for instructional purposes into groups of approximately equal ability.

6) To determine the amount of work a student should be permitted to carry.

7) To select students who should receive scholarship aid.

8) To aid in handling probationary and morally delinquent students.

9) To discover and deal with those who are wilfully slighting college work—"loafers."

10) To dismiss students for poor work.
11) To reinstate students.
12) To recommend students for positions.
13) To constitute a part of a general scheme of educational and vocational advice and direction for freshmen and others.

The results of intelligence tests of college students show that the college student is highly selected from the point of view of intellect as compared with the population in general. He falls well within the upper 50 per cent, and probably largely within the upper 25 per cent. There is nevertheless a large range of intellect among college students in general. The average score on the Army Alpha Intelligence Test varies from an average of 150 for one institution down to an average of 80 for another, among a group of 21 colleges and professional schools reporting such results (54). Table 2 (78) shows the mean scores on the Thorndike Intelligence Examination for various college groups. Evidently there is a large difference between various colleges as to the mental ability of their students.

TABLE 2

	Mean	S.D.	N
Columbia candidates, 1919	79	14.5	505
Columbia freshmen admitted, 1919	81	13.3	350
Brown University freshmen, 1919	76	15.0	300
Pomona College freshmen, men	74	14.5	102
Pomona College, freshmen, women	71	12.2	106
Stanford freshmen, men	70	15.6	634
Stanford freshmen, women	71	13.7	119
Mills College freshmen	69	11.7	185
Collegiate Institute, Canada, graduates	68	13.0	88
University of California, freshmen	68	15.0	157
Southern State Women's College freshmen	66	13.2	100
Trenton Normal School freshmen	64	13.0	170

Similarly, within any one institution, the various colleges may themselves vary as to the intelligence of their student bodies. In a survey of the Ohio State University by means of Army Alpha, the median scores given in Table 3 were reported by Burtt and Arps (9).

The correlation of intelligence-test scores with college marks is not

TABLE 3

College	Median Score	N
University as a whole	136	5950
Graduate School	157	135
Commerce and Journalism	146	212
Law	142	141
Medicine	142	141
Engineering	141	1392
Agriculture	133	859
Education	133	382
Pharmacy	125	109
Dentistry	115	152
Veterinary Medicine	112	93

high. The correlations reported with marks in one subject rarely ever rise to .50, and many are so low as to denote practically no correlation whatever. When, however, the total year's work of a student is pooled, the correlations with intelligence generally rise. With the average intelligence test these correlations cluster around .50; with long and thorough intelligence test such correlations may rise as high as .70. The mere sectioning of college students according to intelligence does not result in better work being done by the more intelligent unless there are changes in the method of instruction (8). In their total college work those who score low on intelligence tests have more difficulty in meeting requirements than those who score high. Erffmeyer (16) found that of those scoring low 38 per cent were conditioned in their work, 64 per cent failed, and 74 per cent were delinquent in their work; while of those scoring high only 3 per cent were conditioned in their work, only 9 per cent failed, and 12 per cent were delinquent.

We see, therefore, that intelligence tests are playing a very definite part in college administration. In most universities they have passed the experimental stage, and are now being used as a recognized aid in the guidance of the student body.

Race Differences. Because of the large number of children of different racial stocks in our schools, there is a continual interest in racial differences in intelligence. The most extensive comparisons have been made between negroes and whites in this country. In terms of I.Q. where the white child makes an average of 100, the negro child makes an average of from 80 to 90. Several workers have pointed out that the average negro I.Q. for grades or ages tends to decrease. Thus Lacy (47) gives the comparison of average I.Q.'s for negroes and whites (Table 4). There is a tendency to interpret this as due to the influence of environment on the child. Other workers, however, do not find this decrease. There is one report of an advance in I.Q. from Grade 3 to Grade 7 (24). In terms of the overlap of one racial group upon the other, it has been customary to say that about 25 per cent of the negroes reach or exceed the median of the whites. This is in all probability a generous estimate. In high school, where one would expect a very highly selected group of negroes, Thorndike (72*a*) found that only four per cent of the colored reached or exceeded the median of the whites. Peterson's (52) extensive comparison of whites and negroes living in the same localities found that 84 per cent of the whites reached or exceeded the median of the negroes. In terms of the probable error, Peterson finds the mean race difference to be 1.4 P.E.

TABLE 4

	Average I.Q.	
	White	Colored
Kindergarten	103	99
Grade 1	104	93
Grade 2	104	88
Grade 3	102	87

Practically all of the many comparisons of negro and white children show marked differences on intelligence tests. These differences are generally interpreted as a real racial difference in intellectual capacity existing between negroes and whites in this country. Some investigators, however, believe that these differences reflect the differences in educational opportunity and social advantages enjoyed by the two racial groups. Which of these interpretations is correct, it is impossible to say. At present there seems more evidence for the former than for the latter. At any rate such differences as exist are not great enough to call for any differential educational treatment of the two racial groups. The overlapping is so great that there is no psychological justification for different schools. If such are thought desirable they must be defended on other grounds than those of intellectual differences between the two races.

Other racial groups in our schools have been compared on intelligence tests. The children entering into the comparisons are for the most part children born in this country of foreign parentage, and the birthplace of the parents is taken as an indication of their racial stock. With reference to most of these groups we are dealing with children whose parents speak a foreign language. The children themselves are frequently bilingual, speaking a foreign language in the home and English in the school. When such children are tested by means of the usual verbal intelligence tests it is maintained by some psychologists that the resulting scores are depressed because of the language handicap which is operating. The evidence for this view is found in the relatively lower ratings on verbal as contrasted with non-language and performance tests. We must keep this in mind in interpreting the results of our intelligence testing.

Children of Italian parentage have been frequently compared with the norms for the country at large. On the Binet test there are many reports from many different sections of the country showing an average I.Q. generally between 85 and 90. On non-language and performance tests the Italians come closer to the American children than they do on verbal tests, but nevertheless they remain below the American norm.

Somewhat the same situation seems to hold of Polish and other Slavonic racial groups, with the notable exception of the Bohemian group which is very near, if not at, the American norm.

English, Scotch, German, and other north European groups, in so far as they have been tested, seem to approximate the American norms very closely. Jewish children are generally reported to be up to and sometimes above the American norms, in spite of any language handicap from which they may suffer. The few Japanese and Chinese groups that have been tested almost always make a good showing. Mexican children test very low, both on verbal and performance tests. A great many American Indian children have been tested. On verbal tests the average I.Q. is generally between 70 and 80. Jamieson and Sandiford (39), however, found the median I.Q.'s for American Indians on different types of tests (Table 5). The first two tests depend upon an understanding of the English language; the latter two do not.

TABLE 5

Verbal	National Intelligence Test	80
Non-verbal	Pintner Cunningham Primary	78
Non-language	Pintner–Non-Language	97
Performance	Pintner-Paterson Scale	92

All of these results are of children in this country and from them we can draw no conclusions as to the mentality of the races from which these children have sprung. The representatives of any race in the United States are a selection of that race and by no means a random sampling. The forces leading to immigration have been different in different countries and at different times, and so we may get a poor sampling from one country and a good sampling from another. The only way to get real comparisons between races is to get representative samplings in their home countries by means of non-language and performance tests. The value of our present results is to give us a picture of the intelligence differences of the racial groups in the population of the United States. They show a great overlapping of groups from the point of view of intelligence. They indicate much inferior intellect in certain recently arrived alien groups and good intellect in others. The present quota law has cut off a great deal of inferior intellect from Europe, but it seems to have attracted a great deal of still more inferior intellect from Mexico. Although intellect can never be the sole determining basis for the admission of the immigrant, it should nevertheless be an important factor.

Handicapped Children. In our educational system there are numerous groups of handicapped children, and intelligence testing has thrown some light on these groups. We have now special classes for the crippled, the speech defective, the cardiac, the neurotic, the hard of hearing, the partially sighted, the truant, the delinquent, the deaf, the blind, and so on. We cannot take up all these different groups in detail, and indeed we know very little about the intelligence of some of them. Some work has been done on the deaf, the blind, and the delinquent, and we shall restrict our account to them.

The deaf. A very large number of deaf children have been given group intelligence tests, so large a number indeed as to represent a very good sampling of children in special schools for the deaf. The difficulty of testing such children is great, because all language tests must be ruled out, if we want to make comparisons between them and hearing children. Hence the only adequate tests are those of the non-language and performance type. Comparisons have been made between deaf and hearing children by Pintner (56). In terms of M.A. on the Pintner Non-Language Test we find that the average 12-year-old deaf child gets an M.A. of 10 in terms of hearing norms; the average 13-year-old deaf an M.A. of 10-6; the average 14-year-old about an M.A. of 11; and the average 15-year-old about an M.A. of 12. If we turn these into I.Q.'s we get 83, 81, 79, and 80 for the four ages. And this perhaps is the best estimate which we possess at the present time of the intellect of the deaf child as compared with the hearing. To the handicap of deafness, therefore, is added the additional load of a lowered intellectual ability. These things make the education of the deaf extremely difficult. Special teachers and special methods of instruction are a necessity.

The blind. With reference to the blind, the work that has been done does not seem to show as large a difference in intelligence between the blind and seeing as between the deaf and hearing. An adaptation of the Binet scale for the blind applied to 670 blind children shows a larger percentage of feeble-minded, borderline, and dull subjects than is the case with unselected seeing children (30). Furthermore, in an extended series of group tests given by Hayes, we note that, wherever adequate comparisons between blind and sighted subjects can be made, the blind are usually inferior at every age. The blind would, therefore, seem to be somewhat retarded intellectually as compared with the sighted, and there is no evidence that compensation for the handicap of blindness exists in the form of increased sensitivity of touch or hearing or of better memory ability.

The delinquent. The delinquent child is not only a school problem but a serious social problem and a great number of studies have been made with reference to his intellectual status. Pintner (54) in 1923 gave a list of 31 studies dealing with intelligence tests given to delinquent children, and pointed out that the earlier studies gave a median percentage of 64 per cent feeble-minded while the later studies showed only 26 per cent feeble-minded. Since that time the tendency to be conservative has increased. Healy and Bronner's (31) extended study of 4000 repeated delinquents designates only 13.5 per cent as clearly feeble-minded. Their distribution of the I.Q.'s of delinquent children shows the median I.Q. to be about 90. Hence we may say that among delinquent children we find all degrees of intelligence, but the tendency for them as a group is to be somewhat below normal and, furthermore, a disproportionate number of feeble-minded children seem to become delinquent. A child does not become delinquent simply and solely because of lack of intelligence. Lack of intelligence is only one of the numerous factors associated with delinquent conduct. In a careful comparison of delinquent and non-delinquent children living in the same environment, Burt (7) found the following conditions to be closely associated with delinquent conduct: defective discipline, general emotional instability, morbid emotional conditions, a family history of vice and crime, intellectual disabilities such as backwardness and dullness, and so forth. Intelligence testing is of value in the problem of dealing with the delinquent child inasmuch as it will enable us to understand him better and to provide for him a type of education adjusted to his intellectual needs. In helping to bring about a better adjustment of the child to the school, intelligence tests are to that extent an aid in the complicated problem of delinquency.

The Concept of General Intelligence

We have given an account of the tools at the disposal of the psychologist for the measurement of intelligence and also a survey of the uses and results of the intelligence testing that he has undertaken. And we have done this without asking what the psychologist means by intelligence, because the meaning becomes clearer through the results of the work. The measurement of intelligence has grown in a practical way without waiting for a precise definition of this "general intelligence" which is being measured. And it is a good thing that the practical work did not wait for a precise meaning of the term, for the practical results of the measurement of intelligence have helped to put meaning into the concept of general intelligence. To be sure, from the very beginning we find Binet struggling with a definition of intelligence. Judgment is a term Binet frequently uses in his work. Intelligence, he says, "is judgment or common sense, initiative, the ability to adapt oneself." And again he says, "To judge well, understand well, reason well—these are the essentials of intelligence." And again he emphasizes three points: (1) the ability to take and maintain a given mental set; (2) the capacity to make adaptations for the purpose of attaining a desired end; (3) the power of auto-criticism.

Definitions. Other definitions have been contributed by numerous writers. Some of these are:

Intelligence is the ability to adjust oneself to new situations, making the proper use of one's thinking capacity (Stern).

It is the power of readjustment to relatively novel situations by organizing new psychophysical combinations (Burt).

It is a group of innate capacities by virtue of which the individual is capable of learning in a greater or less degree in terms of the amount of these innate capacities with which he is endowed (Colvin).

An individual is intelligent in proportion as he is able to carry on abstract thinking (Terman).

Intelligence involves two factors—the capacity for knowledge and the knowledge possessed (Henmon).

Intelligence seems to be a biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior (Peterson).

Thorndike's View. Thorndike's (75) definition of intelligence is: "The power of good responses from the point of view of truth or fact." And again, speaking of measures of intelligence, he says, "they are measures of certain abilities which contribute to, or accompany, or indicate the existence of, the ability to get and use the truth." His direct practical approach to the problem is again seen in what he calls his first approximation: "For a first approximation, let intellect be defined as that quality of mind (or brain or behavior if one prefers) in respect to which Aristotle, Plato, Thucydides, and the like, differed most from Athenian idiots of their day, or in respect to which the lawyers, physicians, scientists, scholars, and editors of reputed greatest ability at constant age, say a dozen of each, differ most from idiots of that age in our asylums."

All intelligence tests are inventories. They show what intellectual tasks have been accomplished by the individual tested. If the tasks are arranged in levels of difficulty, then one can measure the altitude, range, and area of intellect. "Altitude is the degree of difficulty at which a given percentage of success is attained." "Width or range is the percentage of success at any given altitude or the average percentage of success at any given series of altitudes." "Area is the total number of tasks done correctly, or the percentage which this total is of the number of tasks in the entire list."

Any specific intelligence test measures only a certain segment of intellect, and it is important that we should recognize it as such. Thorndike, therefore, refers to intellect CAVD as that segment of intellect measured by the CAVD test, that is, a test for completion, arithmetic, vocabulary, and directions. By the addition of any other type of task, we could extend the segment of intellect measured. General intelligence would presumably be measured by taking all types of intellectual tasks. In actual practice, this is impossible, hence we must infer the "general intelligence" of an individual from the few measured segments of intellect.

As to the nature of intellect in general, Thorndike says, "The hypothesis that we present and shall defend admits the distinction in respect of surface behavior, but asserts that in their deeper nature the higher forms of intellectual operation are identical with mere association or connection forming, depending upon the same sort of physiological connections, but requiring many more of them. By the same argument the person whose intellect is greater or higher or better than that of another person differs from him in the last analysis in having, not a new sort of physiological process, but simply a larger number of connections of the ordinary sort." There are no qualitative differences in intellect as an original capacity. The quality of the product of intellect depends upon the quantity or number of connections or *C*'s.

Spearman's View. Spearman (64) finds that intelligence is mainly concerned with the eduction of relations and eduction of correlates, which are the main processes of all new cognition. And, furthermore, "All branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements seem in every case to be wholly different from that in all the others." This theory of intelligence is frequently called the two-factor theory, because of the two factors which enter into any intellectual act. The general factor is *"g"* and represents general ability or general intelligence. It is universally present in all intellectual tasks and is constant for the same individual. It is the thing we try to measure when we speak of an individual's native ability. The specific factor is *"s,"* and it represents the specific ability peculiar to the task in question, such as language ability or musical ability or arithmetical ability. The *s* will vary in amount from task to task within the same individual. An individual's score on an intelligence test is due to these two factors, his general ability plus his special ability for that type of test.

The *g* is psychologically general ability. Physiologically it may be general nervous energy. It is some force capable of being transferred from one mental operation to another different one. At present the statement that g is mental energy is a hypothesis, but "there seem to be grounds for hoping that a material energy of the kind required by psychologists will some day actually be discovered."

Theory and Practice. In actual practice these theories and definitions have not influenced test-makers very much. The testing of intelligence began and developed without guidance from any particular theory. One might say that the concept of general intelligence arose as a result of the tests. Theories have followed results, except in the case of Spearman's theory, which was well-formulated early in the history of mental testing.

In the practical construction of tests of intelligence, items are chosen that correlate with some outside criterion of intelligence. Among school children, ability to get along in school has been taken as an indication of general intelligence, and the greater number of tests have been constructed with this as a criterion. Hence the greater number of our tests are scholastic in nature. They are predominantly verbal. The fact that this type of test works best with school children has led some psychologists to identify intelligence with verbal ability.

In contrast to this limitation of the meaning of intelligence, we have the view that intelligence may show itself in any type of reaction. Intelligence is always an evaluation of the reactions of the individual and these reactions may be to any stimulus in our environment. From this view comes the emphasis upon performance tests, mechanical tests, nonverbal, and non-language tests. There may, therefore, be many different kinds of intelligence, such as scholastic intelligence, abstract intelligence, verbal intelligence, concrete intelligence, mechanical intelligence, social intelligence, and the like. General intelligence would then be an average or median of adequate measures of all these specific kinds. This view of intelligence is not a theory of intelligence, because the adequate, good, or satisfactory reactions, which we evaluate as intelligent, may be due to the number of connections in the nervous system, as Thorndike maintains, or again may be due to the central fund of nervous energy suggested by Spearman. Both of these theories seek to explain how intelligent reactions take place, or rather how one type of reaction, which we call intelligent, differs from another type of reaction, which we call unintelligent. Both theories go back to the functioning of the nervous system for an explanation, and we may have to wait for further research of the physiologists and anatomists to help us to the final explanation.

In spite of these difficulties as to the real nature of intelligence, the practical work of testing, has, as we have seen, gone ahead vigorously and has increased our knowledge as to the distribution of intelligence, the growth of intelligence, the regularity of this growth, and other allied problems, and it is to these questions that we shall now turn.

Distribution of Intelligence. It used to be thought that great intellects, such as geniuses, were widely separated from the mass of mankind. And again, those of low intellect, such as the idiots and imbeciles, were considered a separate species. Average or normal people stood in between these

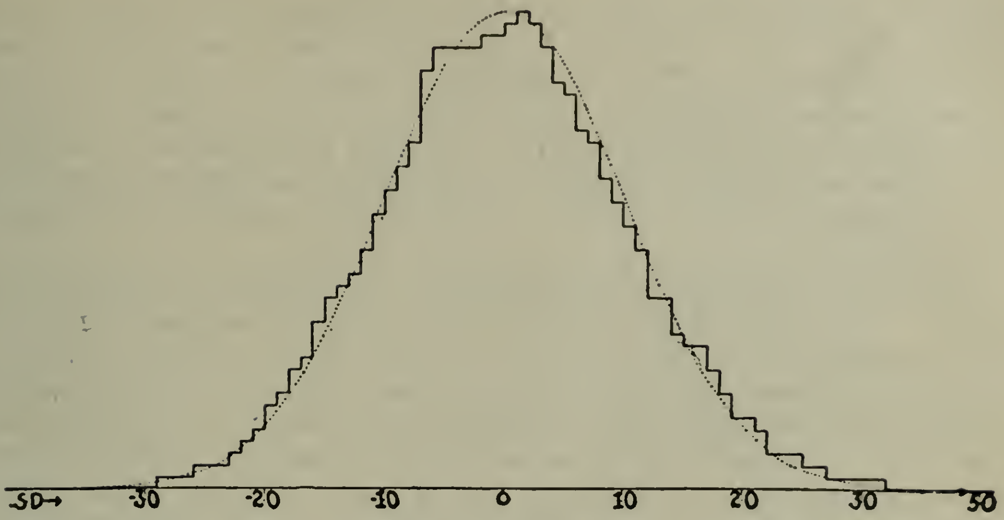


FIGURE 4

Distribution of Intelligence Scores in the Sixth Grade, Based upon Eleven Single Curves

The broken line indicates the theoretical normal curve. (From: Thorndike, E. L. The measurement of intelligence. New York: Teach. Coll. Bur. Publs., 1927. P. 527.)

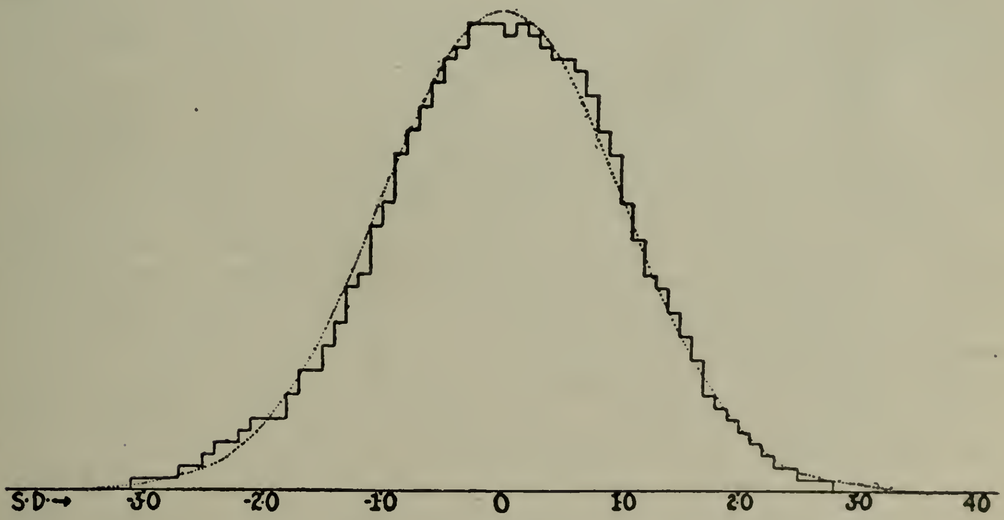


FIGURE 5

Distribution of Intelligence Scores in the Ninth Grade, Based upon Single Curves for the Same Eleven Tests from which the Sixth Grade Composite (Figure 4) Was Derived

The broken line indicates the theoretical normal curve. (From: Thorndike, E. L. The measurement of intelligence. New York: Teach. Coll. Bur. Publs., 1927. P. 529.)

two extreme groups, and they were thought all to have about the same amount of intellect, differences in achievement being due to differences in effort or opportunity. So it was considered the proper and desirable thing for the schoolmaster to flog all who did not learn their lessons, that is, all except the unmistakable idiots or imbeciles, who were excused from school and allowed special privileges. This tripartite arrangement of mankind with reference to intellect formed the underlying assumption of much of our educational practice until very recently, and the influence of this belief is still with us.

The results of intelligence testing show a very different arrangement of intellect. The distribution of scores for any large unselected group is like a so-called normal curve, that is, a bell-shaped curve having few cases at the two extremes with more and more cases as we approach the middle. In terms of an intelligence test this means that there are a few people who make very low scores and that, as the scores increase, the number of individuals making such scores also increases up to a given point. This highest point on the curve represents average or normal ability, because a great number of individuals cluster around this point. From this point on, as the scores increase, the number of individuals becomes less and less until we end up with few individuals making extremely high scores. The frequency of very low and very high scores is small. There are few idiots and few geniuses in the world. The frequency of the intermediate levels is somewhat greater. There are more dullards and bright people. Most numerous of all are the people of normal intelligence, who cannot be called dull or bright and who are far from being idiots or geniuses. Examples of such curves for the sixth, ninth, and twelfth grades, taken from Thorndike (75), are shown in Figures 4, 5, and 6.

The Growth of Intelligence. Intelligence tests have shed some light upon the growth of intelligence from infancy to maturity. There is still much dispute as to the actual shape of the curve. The lack of a scale of intelligence of equal units and the necessity for testing the same group of children from year to year has made a definite answer to this problem difficult. The most common interpretation of such facts as we have leads to a curve of the type shown in Figure 7. If this is a true picture of the growth of intelligence, it means that the increments of growth become smaller and smaller from infancy to maturity. This would seem to fit in with common observation. A 12-months' growth in the infant seems to be much greater than a similar period of growth in the child, and the latter during 12 months seems to grow much faster than the adolescent.

There are psychologists, however, who are doubtful about the shape of the curve pictured in Figure 7. Thurstone argues from a study of the results of the Binet Test for a curve of growth that begins at infancy with small increments, increasing gradually to larger and larger increments up to about age 10 or 12 and thence continuing with decreasing increments until maturity. Freeman (18) from a study of the age curves of many different tests argues for a straight-line curve, at least during the school

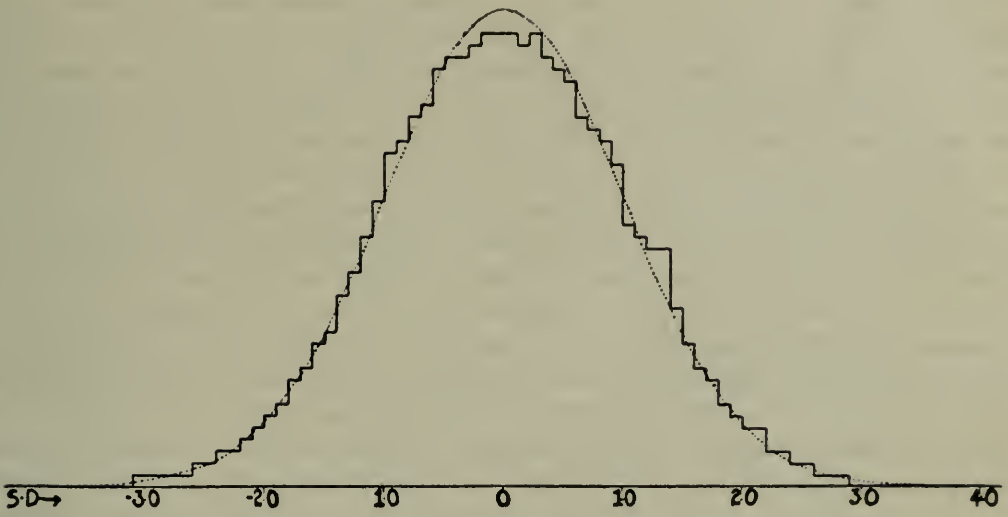


FIGURE 6

Composite Curve for the Twelfth Grade, Derived from Ten Single Curves
The broken line indicates the theoretical normal curve. (From: Thorndike, E. L. The measurement of intelligence. New York: Teach. Coll. Bur. Publs., 1927. P. 539.)

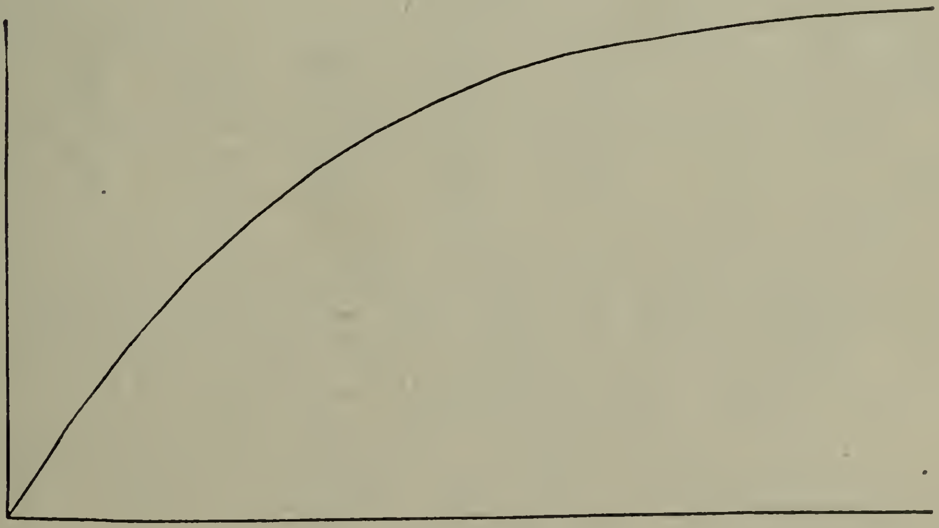


FIGURE 7

The General Nature of the Relation of Altitude of Intellect to Age in Years from 0 to 20

(From Thorndike, E. L. The measurement of intelligence. New York: Teach. Coll. Bur. Publs., 1927. P. 466.)

period, which would mean constant increment of growth from year to year.

The Limit of Growth. Another difficult problem which arises is the determination of the point at which intelligence stops growing. Our curve in Figure 7 shows that somewhere or other the curve becomes flat, the increment of growth becomes zero. If we study the actual age curves for many tests, we may find this upper limit at any age, depending upon the difficulty of the test and upon the nature of the selection of individuals at the different ages. We can take a well-constructed test like the Terman Group Test and plot the scores by age groups for the elementary school and find that the curve begins to flatten out at age 13 or 14; if we add to these the scores of high school pupils, we can make the curve continue upward to age 17 or 18; and again, if we add the scores of college students, we can make the curve continue upward beyond age 17 or 18. Our difficulty lies, of course, with the differences in the selection of cases at different ages. Our elementary school children are least selected from the point of view of intellect, our high-school cases are more highly selected, and our college cases are very highly selected indeed. Unless, therefore, we are very cautious as to the selection of our various age groups, we may unwittingly exaggerate the amount of increase in intellect very greatly.

The best attempt to avoid selective factors and include all the population of a community has recently been made by H. E. Jones whose results are quoted by Thorndike (73). The median scores on Army Alpha by age groups with the number of cases are given in Table 6.

TABLE 6

Age group	11–12	13–14	15–16	17–18	19–21	22–24
Median	51	66	89	93	94	90
No. of cases	125	129	137	98	88	44
Age group	25–29	30–34	35–39	40–44	45–49	50–54
Median	86	83	84	90	73	73
No. of cases	90	104	94	95	69	56

If we take the scores as they stand, we should have to say that intellect increases rapidly from age 11 to age 15-16; from then on it increases very slowly to age 20, where it reaches its highest point; from this point it decreases gradually to age 30; it then remains constant until the early forties where it takes a little upward spurt, only to fall down rather decidely in the late forties and fifties. The upward spurt in the early forties is obviously artificial. Is the more definite spurt between 17 and 24 also artificial or is there a real increase here followed by a decrease? The more extensive data on army officers (81, Table 369) show no such decided decrease in score from age 20 to 30, neither do the scant data on enlisted men (81, pp. 817-818). If, therefore, we are to accept the Jones results at their face value, we must assume a decrease in intelligence immediately after the high point has been reached at about age 20. If we do not assume this, we must suppose that the high scores of those between ages 17 and 24 are due to some disturbing element, and that the curve flattens down rapidly somewhere about 14 or 15, merging into the flat part of the curve from 25 to 40.

This last supposition brings us back to the vast mass of evidence from other tests. Evidently growth in intelligence is fairly vigorous up to age 14 to 16; from then on it is very slow and just where it ceases we are uncertain. It does not sound reasonable to imagine that growth ceases around age 14, and yet much of our evidence points that way. We have some suggestions that it may go on up to about age 20. The crucial zone is the period between 14 and 20. We must wait for further evidence to show us the precise shape of the curve during this period.

The reason for this great interest on the part of psychologists as to the limit of intellectual growth arises because of the necessity for some decision when calculating I.Q.'s. It is the I.Q. technique when applied to older children that has forced the worker to postulate an upper limit of mental growth. If this practical necessity did not exist, there would not be such a pressure to find this upper limit. We cannot go on using the C.A. for older children, because that would yield a diminishing I.Q. Hence there are two problems in regard to this upper limit of mental growth, the theoretical one as to the real shape of the growth curve, and the practical one as to what denominator we are to use for older children in the calculation of the I.Q. The theoretical question is, as we have said, still unsettled; the practical one depends on the most reasonable results which we may get with our present tests until the theoretical problem is solved. A very common procedure is to postulate age 16 as the upper limit of marked growth for our present tests and to use 16 as the denominator in the calculation of I.Q.'s for all children aged 16 and above. There are some psychologists, however, who recommend 14, and it would seem to the writer that 14 is at present a better value to use than 16. This is merely based upon the actual results obtained. Age 14 does not show a decreasing I.Q. in high-school grades as we go from the ninth to the twelfth grade, nor does it result in such a large percentage of low I.Q.'s with adolescents, such as we find in reform schools and the like. Age 14 seems to give, in actual practice, a better distribution of I.Q.'s for these older children than does age 16 with our present tests with their existing standardizations. This use of 14 at the present time is merely a practical expedient for those who wish to use I.Q.'s. It is in no sense a proof of the cessation of mental growth at age 14, and it should not be so interpreted by the worker in mental tests.

The Regularity of Growth. Our growth curve in Figure 7 is a smoothed curve representing the average growth of children in general. Are the separate growth curves for individuals also smooth and regular? Does growth proceed regularly from year to year? In other words, is the I.Q. constant from year to year?

We possess few growth curves for individuals tested from year to year, but, such as we have, they suggest a fairly constant growth. On the other hand, we possess a great number of studies on the constancy of the I.Q.

based upon retests of large numbers of children. A very common way of indicating the amount of constancy or agreement of one I.Q. with a later I.Q. is by means of correlation. The range of 26 such correlations reported by various workers at various times for the Binet tests is from .69 to .98, with a median coefficient of .85. This correlation of .85 probably represents fairly well the amount of agreement we may expect to find between one I.Q. and a second I.Q. for the same child at a later date when tested by an instrument as reliable as the Binet Test. And this is what the psychologist means when he says that the I.Q. is constant. He does not mean that the I.Q. never varies from one examination to another. He simply means that in general the I.Q. tends to remain constant to the amount expressed by a correlation of .85. He means that one I.Q. tends to agree with a second I.Q. of the same individual, rather than to have no relation with it, as would be expressed by a correlation of zero. If the I.Q. were inconstant, the correlation would be near zero, but as the correlation is positive and high we call the I.Q. constant.

A correlation of .85, however, shows that there may be, in any one individual's case, considerable fluctuation from test to test, and such, of course, is the case. Extrinsic factors will account for some of this fluctuation, such as differences between examiners, unreliability of the test, coarseness of the scoring units of the tests, differences in cooperation of the child due to nervousness, fear, slight illnesses, and the like. But even if all such extrinsic factors are avoided so far as possible, there still would appear to be individual fluctuations, and there seems no theoretical reason why there should not be such. There may be, and there very likely are, irregularities in growth from year to year in the mental development of the individual child, just as there are such irregularities in physical growth. But on the whole these are few, and in general the picture of mental growth is that of a fairly regular increase from year to year. With better tests and ideal conditions for examining, the chances are that the constancy of the I.Q. will increase above .85, but there is no reason at all to assume that it will ever reach unity.

Practice and Coaching. An intelligence test is a sampling of stimuli presented to the subject, and to these he reacts. An evaluation of his reactions in the light of what other subjects have done leads to an intelligence rating. If these stimuli are repeatedly presented to the subject, the laws of learning will, of course, operate, and his reactions will tend to improve in accordance with the laws of exercise and effect. There is nothing magical about the material of an intelligence test that puts it beyond the influence of practice. Hence we find that repeated practice of an intelligence test tends to improve the score. If the practice is concentrated and thorough, the scores will increase rapidly and, of course, we will not be able to evaluate such scores in terms of intelligence by comparing them with a standardization group which has had no specific practice. In general, therefore, where we wish to evaluate scores in terms of intelligence ratings by the usual procedure of comparison with a standardization group, we must guard against specific practice on the test to be used.

The Stanford Revision of the Binet would seem not to be greatly influenced by such practice as a child may get by taking the test once each year or even once every six months. But more frequent testings may reveal some effect of practice and so introduce an error into our ratings. Constant use of the Stanford Revision in a school building may result in a degree of familiarity with the tests, through conversation among children, as to slightly increase all the intelligence ratings. In such a case it is well to drop the test for a year or so and substitute another similar scale. In general the same comments apply to group tests. Here the situation is somewhat simpler as we have a greater number of similar group tests to work with and there is no need to use any one in a school to the extent of pronounced familiarity.

More difficult to evaluate and to make allowance for is the general improvement that would seem to come from familiarity with tests in general. If children are constantly taking tests, they would seem to increase their scores slightly above those of children not familiar with tests. How much this affects our results, it is difficult to say. The effect will not be as great as that resulting from practice on a given test. In all probability the effect from familiarity with tests in general is relatively slight.

Very different is the situation with respect to deliberate coaching on any given test. This can very markedly increase the score. Experimental work (28) has shown that by definite teaching of the items of the Stanford Revision the mental ages of children can be increased about two years. After a year, however, the effects of such coaching seem to have largely disappeared. The score of a child who has been definitely coached on a test cannot therefore be evaluated. In actual practice it is generally fairly easy to discover the child who has been deliberately coached on an individual test. He usually gives himself away some time during the examination. It is not so easy to discover the coached cases on a group test.

General Environmental Influences. How much differences in environment, apart from specific training or practice on the tests themselves, influence the scores on our intelligence tests has been very much debated. At the one extreme there seem to be workers who regard the tests as more or less magical measures impervious to any influences of environment. At the other extreme there are those who would seem to suppose that any slight change in the environment of a child will increase or depress his score. If we keep clearly in mind just what intelligence tests are, we may arrive at a better understanding of the problem. An intelligence test offers a sampling of reactions to which the child responds. Then these responses of the child are compared with a great number of responses given to these same stimuli by other children, and in this way the child's responses are rated as good, medium, or poor. The assumption underlying this procedure is that the background of the child to be rated, i.e., his environment and training, is similar to the background of the group with which he is compared. The more these two backgrounds deviate from each other, the more the background of the child to be compared deviates from the background of the comparison group, the less valid become our intelligence ratings. The background of a Russian child differs so much from that of an American child that no one would dream of giving the National Intelligence Test in the same way (i.e., in English) to the two children and then comparing them as to intelligence. Such a procedure is obviously absurd. Let us now come one step nearer. Is the background of the children of immigrants in a large New York public school similar enough to that of American children in general to allow us to get valid intelligence ratings on the National Intelligence Test? Some psychologists doubt this very much and believe that the ratings in such a case will be all somewhat depressed.

The Binet Revision assumes a school environment from about age six onwards. It has been standardized on school children. Gordon (27) in England found that canal-boat children showed a lower and lower I.Q. as they grew older. At ages six or seven, when compared with the standards of the Binet Scale, the intelligence ratings are probably fairly valid. The environmental differences between the canal-boat children and the standardization group are probably not sufficiently great to invalidate comparisons. After these ages, however, the standardization group becomes a purely school group, whereas the canal-boat children attend school less and less as they grow up. If tested on the Binet their I.Q.'s go down, but we cannot say that because of the influence of environment they become more and more stupid. It is our intelligence scale that is becoming a less and less adequate instrument of measurement for these cases. Other experimenters (13, 32) have not found any influence on the I.Q. caused by such absence from school as is found among urban children. Evidently our usual group tests and Binet tests are valid instruments for school children in general regardless of such iregularity in schooling as normally occurs.

Most of the discussion as to the influence of environment on intelligence test scores has centered round the Binet Scale. Two recent studies, one by Freeman (19) and the other by Burks (6), have attempted to measure what change in the intelligence rating on the Binet Scale could be brought about by radical changes in the environment of a child. These studies deal with the changes produced in dependent children adopted into foster homes, and the comparison of the resemblance in intelligence between true children and true parents as contrasted with that between foster children and foster parents. Both studies show the influence of environmental changes on the intelligence rating. Freeman computes the average gain in I.Q. due to the influence of the foster home as 7.5 points; the gain made by those placed in the best homes is estimated at 10.4 points. Burks estimates the average gain to be about 5 or 6 points. We must not, therefore, think of a Binet I.Q. as a fixed unchangeable measure of a child's intelligence. If a child is taken from a very inferior environment and placed in a very superior environment at a very early age, it would seem possible to expect an increase in I.Q. from 10 to 20 points. Such changes are, however, rare. They are more likely to take place if the child is moved to the good environment when very young and less likely to take place with children of school age.

Freeman tested a certain number of cases on a non-verbal test (The Princeton International Test), in addition to the Binet test. Here we have no age norms, so that I.Q. comparisons are not possible. Comparing the scores of 76 matched pairs, he finds a difference in score of 10.8±3.4 in favor of the siblings in the better homes. This difference, he writes, "is, however, not so significant a difference as that found in the case of the Binet Test. This lack of agreement in degree might be taken to mean that the type of intelligence measured by the Binet Scale is more susceptible to environment, chiefly because of its verbal character." If this is so, further work with non-verbal and non-language tests may show smaller differences in intelligence due to environmental shifts. For the present, however, we must recognize that our intelligence measures are slightly influenced by large shifts in the environment of the child. This does not at all invalidate them for use in school, because very few children are subject to such great changes in environment.

The Inheritance of Intelligence. What evidence is there that intelligence is inherited? Here we find two main lines of investigation: (1) measurement of resemblance by means of correlation; (2) studies of family history and social status. Both of these methods were started by Galton (20-23).

Correlation. Studies of this type show the amount of resemblance existing between pairs of related individuals as contrasted with that between random pairs of unrelated individuals. Unrelated individuals will show no correlation, just because by chance any amount of intelligence may occur in each individual of the random pair. If, however, there are factors at work, such as heredity or similar home training, that make for the resemblance of intelligence between siblings, then the correlation between siblings should be greater than zero. This positive correlation does not prove the heredity of intelligence, because it might be caused by the similarity of environment and training to be found among siblings. If, however, the correlations for intelligence resemble the correlations for physical traits, not subject to environmental influences, then there is a strong suggestion that heredity is a causal factor in the mental as well as the physical traits. The correlations between siblings for such physical traits as color of eyes, color of hair, cephalic index, and the like cluster around .5. Let us see what are the correlations for intelligence.

There are a great number of studies showing the correlations between siblings for all sorts of intellectual performances. The earlier studies dealing with short and very specific tests, such as cancellation, memory, and the like, show correlations clustering around .3. The later studies dealing with what we now call intelligence tests show correlations ranging from .33 to .68. The median correlation of 12 such reports, totalling in all about 5000 pairs of siblings, is .51. This is what we find for physical traits. The resemblance of twins in intelligence has been studied by many workers. The average of several correlations, totalling in all about 400 pairs of twins, is .78. A few correlations have been reported for cousins giving a correlation of about .23.

If now we look at these three degrees of relationship, twins, siblings, cousins, we note that the resemblance decreases as the relationship becomes less close, being about .7 for twins, .5 for siblings, and .2 for cousins. For unrelated children our correlation is zero. Although this does not in any sense prove the inheritance of intelligence, if fits in very decidedly with the assumption that intelligence is inherited in exactly the same way as are physical characteristics.

Family histories. Investigations of families through several generations sometimes show the predominance of certain characteristics. Such predominance may be due to the inheritance of underlying basic abilities or tendencies to reaction. Galton showed the concentration of eminent individuals within his own family group, and Pearson (51) has still further elaborated the genealogical survey of the Galton-Darwin-Wedgewood families. There are many such family histories showing a large number of very intelligent individuals all closely related. And, on the other hand, there are family histories showing great lack of intellect in all the members. One of the earliest of such studies in this country was that of the now famous Jukes family (15). Since then there have been many others. Notable among these is the study of the Kallikak family by Goddard (25*a*), in which we have a legitimate line showing mostly men of high intelligence such as doctors, lawyers, educators, judges, and business men, and also an illegitimate line, springing from a feeble-minded girl, made up mostly of people of very low intelligence.

It has been argued that such family histories do not prove the inheritance of intelligence, because children in the good families have a good environment, and, therefore, tend to be like their forbears, and the children of poor familes suffer from their lack of opportunity. But, in general, the family of sound intellect is conspicuous by its ability to overcome environmental disabilities, whereas the opposite type tends to succumb to them. Estabrook (17) in 1915 made a follow-up study of the Jukes family and found them showing the same characteristics as they had shown in 1877, and, furthermore, migration of some branches of the family westward did not seem to have changed their original native tendencies. None of this, however, can be called rigid proof of the inheritance of general intelligence, but all of it fits in clearly with the theory that general intelligence and other mental traits are inherited.

Relatives of superior children. Further supporting evidence for such a theory is to be found in the great number of superior individuals found in the families of superior children. Terman (72) found that 22.6 per cent of the members of the Hall of Fame were related to one or more of the superior children studied. Twelve of the 643 children were found to have parents or grandparents in *Who's Who in America,* with an average age of fathers of 41 years, as contrasted with the average age of first inclusion in *Who's Who* of 49.9 years. In addition there were 35 other relatives in *Who's Who* and 58 other relatives named in standard cyclopedias of biography.

Social status. The distribution of the occupations of the fathers of

Terman's gifted group compared with the occupations of the general population in the same area is given in Table 7.

TABLE 7

	Percentages	
	Fathers of gifted	General population
Professional	29.1	2.9
Public service	4.5	3.3
Commercial	46.2	36.1
Industrial	20.2	57.7

We have also many comparisons of the intelligence ratings of children of different occupational groups. Sandiford (61) gives the results for about 5000 high-school and college students; Haggerty and Nash (29) for about 8000 children in the grades (Table 8).

TABLE 8

Occupation of fathers	Median I.Q. of children	
	High school	All grades
Professional	105	116
Business	103	107
Skilled	103	98
Semi-skilled	102	95
Farmers	102	91
Unskilled	101	89

The medians decrease in the same order, although they are in general higher for the high school than for the grades. Even in the high school, however, the medians for the lower groupings are somewhat lower than those of the higher groupings in spite of the much more highly selected groups of children in the lower groupings.

There are many other similar studies, all pointing in the same direction. Of course this relationship between the occupation of the parent and the intelligence of the child is no proof of heredity, but taken along with the other facts we have mentioned, it fits in admirably. The more intelligent children tend on the whole to have the more intelligent parents, and such parents tend to be found in greater numbers in the higher occupations. The higher the occupation, the more likely is it to require intelligence for successful adjustment.

SUMMARY

1) Individual differences in intelligence among children have always been vaguely recognized. Intelligence tests attempt to measure these differences more accurately.

2) The early mental tests were largely measurements of separate simple mental and psychophysical functions.

3) Binet combined a great number of simple tests in order to arrive at a measure of general ability. He constructed the first scale with mental-age levels.

4) At present the psychologist has a great number of scales and tests for the measurement of intelligence. They can be conveniently divided into individual and group tests.

5) The individual tests include many scales of the Binet type as well as scales of performance tests, or mixed performance and verbal tests.

6) Group tests are available for all levels of intelligence from the kindergarten level to the graduate-student level. They are largely verbal in content. Non-verbal tests are used more frequently at the lower levels. There are a few non-language tests which can be given without the use of any language.

7) A very common method of evaluating the scores of an intelligence test is to turn them into mental ages, from which intelligence quotients can be calculated. This technique is useful in the grades, but has distinct limitations for older people. An evaluation of scores in terms of the standard deviation is a sounder method.

8) The most frequent use of intelligence tests in school is to aid in the proper classification of children. The mental age helps to determine the grade, while the I.Q. helps to determine the section.

9) Tests are useful in selecting the children that should be placed in special classes for the dull or for the bright.

10) Intelligence scores are one of the many items necessary in educational and vocational guidance.

11) The use of intelligence tests in universities is very widespread. Their results have given us much information about the college student, and the tests are used in many ways by the administrative officers.

12) Tests of different racial groups in the United States indicate slight differences in the average intelligence of the groups. The most adequate comparisons are those between negroes and whites. The other comparisons are complicated by language factors. No adequate comparisons between national groups in their own countries have yet been made.

13) Deaf children when compared with hearing children rate much lower on intelligence tests.

14) Blind children would seem to be somewhat poorer than sighted children as far as intelligence is concerned.

15) Among juvenile delinquents in courts and institutions there would appear to be a larger percentage of dull and feeble-minded children than in the general population, but it is important to remember that all degrees of intelligence are found among delinquent children. Poor intelligence is only one of the many factors associated with delinquency.

16) There are a great many definitions of intelligence. Two common ones are: the ability to adjust and the ability to learn.

17) Thorndike defines intellect as the ability to get and use the truth. He describes intelligence by the attributes of level, range, area, and speed. Differences in intelligence are due to differences in the number of connections in the nervous system.

18) Spearman builds up a theory of intelligence. It is frequently called the two-factor theory, because all intellectual operations depend

upon a general factor and a specific factor. The general factor is present in all intellectual tasks.

19) The distribution of intelligence in the population at large conforms very closely to a normal distribution.

20) The growth of intelligence would seem to be rapid at first with gradually diminishing acceleration.

21) The limit of growth of intelligence is still much disputed. It would seem to lie between 14 and 20.

22) Growth seems to be fairly regular or steady from year to year. This gives a fairly constant I.Q.

23) All tests of intelligence are subject to the usual influences of practice.

24) The Binet tests are subject to environmental influences. Shifts from very poor to very superior environments at an early age may lead to increases in I.Q. up to 20 points.

25) The evidence for the inheritance of intelligence rests on the resemblance of siblings ($r=.5$) and twins ($r=.7$), as well as on the study of family histories.

REFERENCES

1. ARTHUR, G. The re-standardization of a point performance scale. *J. Appl. Psychol.,* 1928, **12**, 278-303.
2. BERE, M. A comparative study of the mental capacity of children of foreign parentage. *Teach. Coll. Contrib. Educ.,* 1924, No. 154.
3. BOOK, W. F. The intelligence of high school seniors. New York: Macmillan, 1922.
4. BRIGHAM, C. C. A study of American intelligence. Princeton, N. J.: Princeton, Univ. Press, 1923.
5. BRONNER, A. F., *et al.* A manual of individual mental tests and testing. Boston, Mass.: Little, Brown, 1927.
6. BURKS, B. S. The relative influence of nature and nurture upon mental development. *27th Yrbk. Nat. Soc. Stud. Educ.,* 1928, Pt. 1, 219-316.
7. BURT, C. The young delinquent. New York: Appleton, 1925.
8. BURTT, H. E., *et al.* Efficiency of instruction in unselected sections in elementary psychology compared with that in sections selected on the basis of intelligence tests. *J. Educ. Psychol.,* 1923, **14**, 154-161.
9. ————. The university intelligence tests. *Ohio State Univ. Bull.,* 1922, **27**, No. 5.
10. DEARBORN, W. F. Intelligence tests. Boston, Mass.: Houghton Mifflin, 1928.
11. DEARBORN, W. F., *et al.* A series of form-board and performance tests of intelligence. *Harvard Univ. Monog. Educ.,* 1923.
12. DECROLY, O., & BUYSE, R. La pratique des tests mentaux. Paris: Alcan, 1928.
13. DENWORTH, K. M. The effect of length of school attendance upon mental and educational ages. *27th Yrbk. Nat. Stud. Educ.,* 1928, Pt. II, 67-91.
14. DICKSON, V. E. Mental tests and the classroom teacher. Yonkers, N. Y.: World Book, 1923.
15. DUGDALE, R. L. The Jukes. (4th ed.) New York: Putnam, 1910.
16. ERFFMEYER, C. E. Intelligence tests as an aid in the diagnosis of academic maladies. *School & Soc.,* 1924, **20**, 317-320.

17. Estabrook, A. H. The Jukes in 1915. *Publ. Carnegie Instit. Wash.,* 1916.

18. Freeman, F. N. Mental tests. Boston, Mass.: Houghton Mifflin, 1926.

19. ———. The influence of environment on the intelligence, school achievement and conduct of foster children. *27th Yrbk. Nat. Soc. Stud. Educ.,* 1928, Pt. 103-217.

20. Galton, F. Hereditary genius. New York: Appleton. 1880. (Reprint.)

21. ———. English men of science. New York: Appleton, 1890. (Reprint.)

22. ———. Natural inheritance. London, 1889.

23. ———. Inquries into the human faculty and its development. New York: Macmillan, 1883.

24. Garth, T. R., & Whatley, C. A. The intelligence of southern negro children. *School & Soc.,* 1925, **22**, 501-504.

25. Gesell, A. The mental growth of the pre-school child. New York: Macmillan, 1925.

25*a*. Goddard, H. H. The Kallikak family. New York: Macmillan, 1914.

26. Goodenough, F. L. The measurement of intelligence by drawings. Yonkers, N. Y.: World Book, 1926.

27. Gordon, H. Mental and scholastic tests among retarded children. London: Board of Education, 1923.

28. Graves, K. B. The influence of specialized training on tests of general intelligence. *Teach. Coll. Contrib. Educ.,* 1924, No. 143.

29. Haggerty, M. E., & Nash, H. B. Mental capacity of children and parental occupation. *J. Educ. Psychol.,* 1924, **15**, 559-572.

30. Hayes, S. P. Ten years of psychological research in schools for the blind. *Publ. Pa. Instit. Blind,* 1927, Jan.

31. Healy, W., & Bronner, A. F. Delinquents and criminals. New York: Macmillan, 1926.

32. Heilman, J. D. The relative influence upon educational achievement of some hereditary and environmental factors. *27th Yrbk. Nat. Soc. Stud. Educ.,* 1928, Pt. II, 35-65.

33. Herring, J. P. The Herring revision of the Binet-Simon tests. Yonkers, N. Y.: World Book, 1922.

34. Hirsch, N. D. A study of natio-racial mental differences. *Genet. Psychol. Monog.,* 1926, **1**.

35. Hollingworth, L. S. Gifted children. New York: Macmillan, 1926.

36. ———. The psychology of subnormal children. New York: Macmillan, 1920.

37. Hopkins, L. T. The intelligence of continuation school children. Cambridge, Mass.: Harvard Univ. Press, 1924.

38. Irwin, E. A., & Marks, L. A. Fitting the school to the child. New York: Macmillan, 1924.

39. Jamieson, E., & Sandiford, P. The mental capacity of Southern Ontario Indians. *J. Educ. Psychol.,* 1928, **19**, 313-328.

40. Johnson, B. J. Mental growth of children. New York: Dutton, 1925.

41. Kelley, T. L. Interpretation of educational measurements. Yonkers, N. Y.: World Book, 1927.

42. ———. Statistical method. New York: Macmillan, 1923.

43. ———. The influence of nurture upon native differences. New York: Macmillan, 1926.

44. Kirkpatrick, C. Intelligence and immigration. Baltimore, Md.; Williams & Wilkins, 1926.

45. Kohs, S. C. Intelligence measurement. New York: Macmillan, 1923.

46. Kuhlmann, F. A handbook of mental tests. Baltimore, Md.: Warwick & York, 1922.

47. LACY, L. D. Relative intelligence of white and colored children. *Elem. School J.*, 1925-1926, **25**, 542-546.

48. MACPHAIL, A. H. The intelligence of college students. Baltimore, Md.: Warwick & York, 1924.

49. MCCALL, W. A. How to measure in education. New York: Macmillan, 1922.

50. PAULU, E. M. Diagnostic testing and remedial teaching. Boston, Mass.: Heath, 1924.

51. PEARSON, K. The life, letters and labours of Francis Galton. (2 vols.) London: Cambridge Univ. Press, 1924.

52. PETERSON, J. The comparative abilities of white and negro children. *Comp. Psychol. Monog.*, 1923, **1**, No. 5.

53. ————. Early conceptions and tests of intelligence. Yonkers, N. Y.: World Book, 1925.

54. PINTNER, R. Intelligence testing. New York: Holt, 1923.

55. ————. An empirical view of intelligence. *J. Educ. Psychol.*, 1926, **17**, 608-616.

56. PINTNER, R., DAY, I. S,. & FUSFELD, H. E. A survey of American schools for the deaf. Washington, D. C.: National Research Council, 1928. See Chap. XIV.

57. PINTNER, R., & PATERSON, D. G. A scale of performance tests. New York: Appleton, 1917.

58. PORTEUS, S. D. Studies in mental deviations. *Vineland Tr. School Bull.*, 1922.

59. PRESSEY, S. L., & PRESSEY, L. C. Introduction to the use of standard tests. Yonkers, N. Y.: World Book, 1922.

60. RUCH, G. M., & STODDARD, G. D. Tests and measurements in high school instruction. Yonkers, N. Y.: World Book, 1927.

61. SANDIFORD, P. Paternal occupations and intelligence of offspring. *School & Soc.*, 1926, **23**, 117-119.

62. SHAKOW, D., & KENT, G. H. The Worcester form board series. *Ped. Sem.*, 1925, **32**, 599-611.

63. SPEARMAN, C. The nature of intelligence and the principles of cognition. London: Macmillan, 1923.

64. ————. The abilities of man. New York: Macmillan, 1927.

65. STEDMAN, L. M. Education of gifted children. Yonkers, N. Y.: World Book, 1924.

66. STERN, W. Die Intelligenz der Kinder und Jugendlichen. Leipzig: Barth, 1920.

67. STUTSMAN, R. Performance tests for children of pre-school age. *Genet. Psychol. Monog.*, 1926, **1**.

68. SYMONDS, P. M. Measurement in secondary education. New York: Macmillan, 1927.

69. Symposium: intelligence and its measurement. *J. Educ. Psychol.*, 1921, **12**, 123-147; 195-216.

70. TERMAN, L. M. The measurement of intelligence. Boston, Mass.: Houghton Mifflin, 1916.

71. ————. The intelligence of school children. Boston, Mass.: Houghton Mifflin, 1919.

72. ————. Genetic study of genius. (Vol. I.) Stanford University, Calif.: Stanford Univ. Press, 1925.

72*a*. THORNDIKE, E. L. Intelligence scores of colored pupils in high schools. *School & Soc.*, 1923, **18**, 569-570.

73. THORNDIKE, E. L., *et al.* Adult learning. New York: Macmillan, 1928.

74. ————. The psychology of algebra. New York: Macmillan, 1923.

75. ————. The measurement of intelligence. *Teach. Coll. Bur. Publ.,* 1927.

76. WELLS, F. L. Mental tests in clinical practice. Yonkers, N. Y.: World Book, 1927.

77. WHIPPLE, G. M. Manual of mental and physical tests. Baltimore, Md.: Warwick & York, Vol. I, 1914; Vol. II, 1915.

78. WOOD, B. D. Measurement in higher education. Yonkers, N. Y.: World Book, 1923.

79. WOODROW, H. Brightness and dullness in children. Philadelphia: Lippincott, 1919.

80. YERKES, R. M., & FOSTER, J. C. A point scale for measuring mental ability. Baltimore, Md.: Warwick & York, 1923.

81. YERKES, R. M. [Ed.] *Mem. Nat. Acad. Sci.,* 1921, **15.**

CHAPTER 18

THE INDIVIDUAL IN SCHOOL: II. SPECIAL ABILITIES AND THEIR MEASUREMENT

FRANK N. FREEMAN
University of Chicago

One of the prominent characteristics of modern psychology is the large emphasis which it places upon individual differences. The interest in individual differences finds expression in the investigations which are carried on in the theoretical laboratory as well as in the applications of psychology to practical problems. The study of individual differences has been especially fruitful in its applications to the problems of industry, of penology, and of education. We are concerned particularly in this chapter with its applications to education.

We may distinguish broadly between two general types of individual differences. The first type may be called differences in temperament and the second type differences in ability. Differences in temperament include differences in voluntary behavior, differences in emotion or feeling, and differences in likes and dislikes, or in interests. Differences in attitude, while they may not strictly be included under the head of temperament, are akin to those which have been mentioned. Somewhat more distantly related still are judgments and attitudes on questions of moral conduct. Attempts have been made to measure all of these types of differences in temperament by means of standardized tests.

In contrast to differences in temperament are the differences in ability. These again fall under two heads, namely, differences in general ability and differences in special ability. Differences in general ability or intelligence were treated in the preceding chapter. Differences in special ability constitute the subject of the present chapter.

THE DISTINCTION BETWEEN GENERAL AND SPECIAL ABILITY

The distinction between general and special ability and the meaning of general ability and of special ability seem clear until we try to analyze the concepts which are involved and to measure the differences between individuals in these various types of ability. The meaning of general ability and the nature of the tests of general ability have been discussed at length in the preceding chapter. General ability, as we have seen, is sometimes looked upon as merely the sum of special abilities. General ability, according to this view, would be a mathematical abstraction rather than a strictly psychological entity. On this theory the only real differences which distinguish individuals from each other are differences in special abilities. On the other hand, some psychologists consider general ability to be a distinctly different trait from special abilities, existing apart from and perhaps independently of the special abilities.

Again, there are different conceptions concerning the meaning of special abilities. We have not been able clearly and definitely to identify and to catalogue the various special abilities. Concerning the existence and nature of some special abilities we are fairly clear; concerning others we are not sure how far they may be elementary and how far they may be mixed, how far they may be special abilities, and how far they may be combined with general ability, and, finally, how far they may be genuine abilities or merely the accidental results of training. We may be fairly sure, for example, that ability in pitch discrimination is a special ability. But should we regard musical ability as a special ability, as a composite of special abilities or as a mixture of special abilities and general ability? Is motor dexterity a special ability, a mixture of special abilities, or the product of training? The same question may be asked concerning differences in memory, differences in artistic ability, mathematical ability, language ability, executive ability, mechanical ability, and so on. We shall return to a further discussion of these questions at the close of the chapter.

The Distinction between Special Aptitude and Achievement

Differences in special ability may have either one of two origins. In the first place they may be native or inherited differences. They may be due to inherent differences in the bodily organism — chiefly in the constitution of the central nervous system. These we call differences in special aptitude. On the other hand, the differences may be due partly or wholly to differences in experience or in formal training or education. When we assume that the differences which we measure are due at least in part to training, or when we wish to make no assumption at all concerning their origin, we usually characterize them as differences in achievement. Differences in special aptitude are differences which are assumed to be largely due to native constitution. Differences in achievement are either those which are assumed to be due to education or concerning which no assumption is made.

We must be on our guard against concluding that an ability is native because the test of this ability is dubbed a special aptitude test. Similarly, we cannot take the score on an achievement test at its face value as a measure of the result of education alone. The relation between nature and nurture is an important problem for investigation in the field of special abilities as it is in the field of general ability. We shall revert again to this problem at the close of the chapter. In the meantime we shall use the term special aptitude test to designate those tests which aim or pretend to measure abilities assumed to be chiefly native in their origin, and we shall use the term achievement test to designate those tests which are designed chiefly to measure the results of education.

Origin and Development of Special Aptitude Tests

The term "special aptitude" in the sense in which it is here used is of comparatively recent origin. The term "aptitude" has been used for a long time to designate differences in ability in a general way, but its use in the plural, to designate particularized abilities which are distinguished from general ability and which are assumed to be largely native in their origin, is a comparatively modern one.

The relative emphasis upon the measurement of general ability, on the one hand, and of special abilities on the other hand has varied at different periods in the history of testing. In the years from 1890-1905 practically all of the mental tests which were used were tests of special ability, or special aptitude tests in the sense in which the term is here used. The abilities which were measured were chiefly sensory and motor abilities, and these abilities have been found throughout the history of testing to be largely specialized in nature. An example of these tests is found in the collection of tests given to Columbia students under the direction of Cattell. The list of these tests as reported by Wissler (65) is as follows:

Length and breadth of head
Strength of hand
Fatigue as measured by an instrument called the dynamometer
Acuity of vision
Color vision
Acuity of hearing
Pitch discrimination
Weight discrimination
Discrimination of two points on the skin by the aesthesiometer
Pain sensation
Perception of size
Color preference
Reaction-time
Rate of the perception and reaction as measured by the rapidity of crossing out *a*'s in a text
Rapidity of naming colors
Rate of movement as measured by dotting in one-centimeter squares with a pencil
Accuracy of movement as measured by striking dots with a pencil
Perception of time as measured by the ability to follow rhythm one second after the sound has ceased
Association as measured by free associations to nine words
Imagery as measured by the imagery test of Galton
Memory as measured by four simple memory tests.

Attention was shifted from such tests of special ability to the attempt to measure more general abilities by the work of Spearman (52, 53) and of Binet (3). Spearman thought that he discovered by means of the intercorrelation of tests the existence of a general or a central factor, existing in various proportions in different composite abilities. Binet in the organization of his scale, which was first published in 1905, ignored the child's responses on the individual tests and considered only the total response on all of the tests of the scale taken together. This emphasis upon general ability and its measurement was continued in the American revision of the Binet Scale and in the group intelligence tests which were devised during the World War and have been extensively used down to the present.

The attempt to distinguish the special abilities and to measure them individually continued alongside the attempt to measure general ability.

Healy and Fernald (24), for example, in the series of tests which they used at the Juvenile Court in Chicago, treated the abilities which they measured as separate abilities and did not combine the scores into a composite.

Woolley and Fischer (69), in the tests which they used in the study of working children in Cincinnati, attempted to measure a variety of abilities separately rather than to measure general or composite ability. Other examples of special tests are Seashore's test of musical capacity (50) and Stenquist's Mechanical Aptitude Test (55). The most extensive tests of special ability have been made for the purpose of finding means of determining vocational aptitudes — that is, of predicting success in particular vocations. Perhaps the earliest of these was Münsterberg's test for motormen (42). Tests have been used to measure the aptitude for the performance of simple factory operations, for performing clerical activities, and for meeting the demands of more complex vocations.

Origin and Development of Educational Tests

Educational tests have had an entirely different origin from special ability tests of the type which we have called special aptitude tests. Educational tests have developed from school examinations. Their aim is to get a more accurate measure of the performance of children in the various subjects of study than is attained by means of the ordinary examinations. The analysis of abilities which is represented in educational tests, or which is derived from their results, is not based upon the general psychological classification of abilities or upon the distinctions between abilities which arise from work in the psychological laboratory. Instead it follows the lines of demarkation between the various school subjects or between the aspects of the school subjects which can be clearly distinguished from each other. For example, a writing test is designed to measure those abilities which are required in writing or in learning to write. Those abilities which are required in reading and in learning to read will be measured in a reading test. The same is true of arithmetic tests, history tests, composition tests, and others. In so far as an educational test is a measure of the individual's ability, then, the ability which is measured by a given subject-matter test is the same as is required to learn this particular subject.

The relationship of educational tests to ordinary school examinations is brought out in a review of the early examples of educational tests. The earliest known example of an educational test as distinguished from an ordinary examination is a device used by George Fisher in the Greenwich Hospital School in England. The report of this device was published in 1864. It consists of a scale-book containing graded samples of pupils' performance. The pupils were to be marked by a comparison with the samples of performance given in the scale-book. The following quotation describes the application of this method in the grading of handwriting.

"A book, called the 'Scale-Book,' has been established, which contains the numbers assigned to each degree of proficiency in the various subjects of examination: for instance, if it be required to determine the numerical equivalent corresponding to any specimen of 'writing,' a comparison is made with various standard specimens, which are arranged in this book in order of merit; the highest being represented by the number 1 and the lowest by 5, and the intermediate values by affixing to these numbers the fractions ¼, ½, or ¾. So long as these standard specimens are preserved in the institution, so long will instant numerical values for proficiency in 'writing' be maintained. And since fac-similes can be multiplied without limit, the same principle might be generally adopted" (8).

The refinement of the examination method in this instance obviously consists in substituting an objective standard of judgment for the standard of judgment existing in the mind of the teacher or of the examiner. This objective standard of judgment will serve to make more uniform the judgments of the same individual from time to time or the judgments of different individuals.

The next example of an educational test is furnished in the simple spelling test used by J. M. Rice (47). During the year 1894-1895, Rice drew up two spelling scales. The first was a list of 50 words, the second was a list of sentences containing words to be spelled. These tests were given in large numbers of schools. By using a common list of words Rice was able to compare directly the achievement of the children in the different schools. One of the problems which this comparison called sharply to the attention of educators was the relationship between spelling efficiency and the amount of time which was spent in spelling drills.

The next test was in the field of arithmetic and was devised by C. W. Stone (56) under the supervision of Thorndike. Stone's test consisted of two parts, one containing examples of the fundamentals of calculation, the other containing reasoning problems. The examples were adapted to the level of the sixth grade. In the following year, 1909, S. A. Courtis (12) organized more systematic sets of examples adapted to all the grades. These arithmetic tests, like the spelling test of Rice, consisted of examples of work selected so as to be representative of the work required in the subject.

The next educational test, Thorndike's Writing Scale (59), was of a different sort. Instead of a series of representative examples or tasks, it consisted of a series of standard specimens with which the pupil's writing might be compared. Its method was similar to the method employed by Fisher in constructing his "scale-book."

It is not necessary to multiply the illustrations. Those which have been given reveal the general characteristic of educational tests. This characteristic, common to all such tests, is that they secure samples of products of the pupils' activities in the subjects in question and provide some means of judging and of expressing the degree of excellence of these products. When a pupil is given a standard set of words to spell, his spelling activity is judged and expressed in terms of the number of these words which he spells correctly. When a specimen of his handwriting is secured under standard conditions and scored by a handwriting scale, the excellence of his handwriting is judged and expressed in terms of the scale value of the step of the scale to which it is equal. In one case the pupil's performance is judged in terms of the number of a series of standard tasks he can perform. In the other case it is judged in terms of the quality of his performance, scored on a scale of merit. Other modes of judging the pupil's performance are used in standard tests, as we shall see, such as accuracy, speed, or degree of difficulty of the task, but in all cases the procedure consists in securing a sample of the pupil's performance under standard conditions and then using some standard procedure of judging that performance.

The distinguishing mark of educational tests, as contrasted with ordinary examinations, is involved in the limiting term *standard*. If we know all that is implied in this term we know the characteristics of educational tests. We shall next discuss the ways in which tests are made standard through the process which is called standardization.

General Characteristics of Standardization

Selection of Representative Samples of Performance. It is quite clear that one of the limitations of ordinary examinations is that they ordinarily fail in some measure to give a true picture of the pupil's mastery of the subject as a whole. The samples of performance which are required of the pupil in the examination may not spread evenly over the whole subject. This unevenness may appear, in the first place, from a comparison of the examination with the course as taught by the teacher making the examination and as studied by the pupil taking it. In the second place, it will appear from a comparison of the examination with the course as taught by other teachers of the subject. When examined from this point of view, the discrepancy will be all the more striking.

Courses in a given subject are known to differ widely as taught by various teachers or as presented in various text-books. This difference is taken as a matter of course. Accordingly, it has been common for colleges and universities which lay down specific entrance requirements to specify not only the subject but the particular topics in the subjects which the student seeking admission is expected to have mastered. For this reason, also, a student would ordinarily not be expected to pass an examination which is not based upon the particular course which he has taken. Students preparing for entrance examinations have commonly taken special pains to study the topics which are usually included in the examinations given by the institution to which they desire admittance. Coaching schools have been run for years, the chief stock in trade of which was a special knowledge of the requirements of particular institutions as expressed in their entrance examinations.

The same situation obtains concerning the examinations in elementary schools and in high schools. If J. M. Rice, instead of giving the same set of words to various schools, had gathered together a collection of average percentages with which the pupils in the various schools spelled correctly the words which were assigned to them, the comparisons would not have been worth the paper on which the tables were written. The significance of the various percentages of correct spelling would have depended altogether upon the words which the children were given to spell. The words taught the pupils of the same grade at different schools would have been found to vary enormously in difficulty. But even if the words taught in different schools had been found to be of the same general level of difficulty, the particular words which were taught would have been found to overlap only to a slight extent.

What was true of spelling was also true of the other subjects, and still is true to a large degree. If one were to examine courses of study and examinations in grammar, history, geography, science, or mathematics, one would find a wide range in the topics covered.

The first problem in standardization of examinations, or in the development of a standardized test, is to select the subject-matter which shall be covered by the test. This subject-matter must be so general or universal in its scope that it can be accepted as representing general practice and not the practice of any individual or small group of individuals. Only a test which is based upon such a representative collection of subject-matter can be considered a test in the subject in general, as distinguished from a test in a particular course.

Broadly speaking, there are two chief methods of establishing the subject-matter which is to be included in a standard test. The first method is the less fundamental but the one more commonly used. This method is to summarize current practice and to include in the test those aspects of the subject which are more commonly taught throughout the country. This summary may be based upon an analysis of textbooks in the subject, or an analysis of courses of study, or both. Such a method does not attempt to find any objective sanction for the topics which are included in the test. It does not assume that the test measures what should be taught but only that it measures what is taught. It does not attempt to guide practice, except in so far as general practice may be considered more desirable than are individual deviations from it. It seeks primarily to reflect practice. A test which reflects general practice is useful in those cases where no method has yet been found or applied of determining, on more fundamental grounds, what should be included in the instruction in the subject.

In some cases a more fundamental method is possible. Such a case is well illustrated in spelling. The Ayres Spelling Scale (1), for example, includes not words which are commonly found in spelling books but words which commonly appear in written correspondence or in printed literature. Another analysis, made by Jones (30), gave a collection of words most frequently used by children in their written composition in school. These pioneer investigations have been supplemented by later studies, notably those by Horn (28).

A spelling scale which is based upon such studies will reveal not merely whether a pupil can spell the words which are commonly taught but whether he can spell the words which he needs to know how to spell. It is clear that tests which are based upon such fundamental study as this are useful not merely as means of measuring the pupil's attainments but also as a guide to educational practice. Somewhat similar surveys as the basis for the organization of a test have been made in the field of language usage. These surveys consisted in finding the errors which are most commonly made in oral speech or in written speech and then including these errors in the test. Examples are the surveys of language errors made by Charters (10) and by Wilson (64).

In some cases it may be assumed that all the elements of a given subject should be learned by the pupils. In such cases the problem of test organization is not a problem of selecting from the various elements but rather the problem of including all of them. An example of this sort is found in the fundamentals of arithmetic. Since the number of combinations in arithmetic in the four fundamentals in numbers below ten is limited and comparatively small, it is possible to include all of them in a standard test.

The selection of standard subject-matter or content is one means of determining that the test shall measure the abilities which it ought to measure, either because they are objectively valuable or because they are characteristic of general practice in teaching the subject. In some subjects of instruction the question concerns the qualities which are to be measured rather than the content or subject-matter. In devising a standard test of handwriting, for example, it is necessary to know what qualities of handwriting should be measured or should be chiefly emphasized. Thorndike (59), in constructing his handwriting scale, endeavored to base the rating of the specimens upon the three characteristics of legibility, beauty, and character. Ayres (2), in constructing his handwriting scale, assumed that the fundamental quality to be measured is legibility. Freeman (15), in his analytical scale, distinguished five qualities of merit, uniformity of alignment, uniformity of slant, letter formation, quality of line, and spacing. These may be summed up under the general head of correct form. The variation in the qualities which are actually measured by the three scales is probably not as great as would be suggested by this variation in statement, but it indicates that there is some difference of opinion as to what qualities should be measured in a test of handwriting.

Standardization in the Form of Test Questions. It is a not uncommon experience for teachers to discover that a question has been misunderstood by a large portion of the class. It is not at all an easy matter to frame questions so that they shall mean exactly the same thing to the students as they do to the person framing them. The difficulty of a question depends not merely upon the difficulty of the mental operation which is required to answer the question but also, in many cases, on the difficulty in understanding the meaning of the question itself. Investigations have shown that the difficulty of a set of reasoning problems in arithmetic may be altered greatly by simply changing the form of statement, without in the slightest changing the mathematical operations which are required. In the construction of standard tests, minute and prolonged attention is given to this matter of the form of statement in which the quesions are couched. In addition to this care in the first statement of the question, a preliminary trial is always given in order to discover statements which are ambiguous or unclear as well as to determine the difficulty of the items of the test in the form in which they are actually used. In the case in which a test is given orally to individual pupils or to groups of pupils, great care must likewise be exercised by the examiner to follow the prescribed form.

Standardization of the Method of Administration. Large variations in the pupil's performance may be caused by variations in the manner in which the instructions for the tests are given. In giving a handwriting test, for example, the teacher may instruct the pupils to write carefully or as well as they can. On the other hand, the teacher may instruct the pupils to write as rapidly as they can. A third method is to tell them to to write both well and rapidly and not to emphasize either speed or quality at the expense of the other. In the type of test in which two possible answers are given and the pupils are instructed to choose between them, the directions may be, in case of doubt, to guess, or, on the other hand, the instructions may be to omit an answer when in doubt. The pupils will react differently under these two types of instruction. A standard test gives comparable results when administered to different groups of pupils by different examiners in part because the details of administration of the test are carefully worked out and are laid down in the form of printed instructions.

"Objectivity" of Scoring. One of the most important causes of variations in the marking of pupils by means of ordinary examinations, of course, is the fact that the same answer may be rated differently by different examiners. Investigations have shown that the variations in the marks assigned by different teachers of the same subject to the same paper may be enormous. Starch and Elliot (54), for example, had the same geometry paper graded by 116 teachers of the subject. Two teachers graded the paper above 90; 20 graded it 80 or above; 47 assigned it a mark above passing or a passing mark; 69 assigned it a failing mark; 20 gave it a mark below 60; and one gave it a mark below 30.

There are two important causes of such variations in examination marks. The first is the variation in the estimate of the performance of the individual pupil in comparison with the performance of other pupils. This variation may be due to the relative emphasis placed by different teachers upon different aspects of the pupil's performance. For example, in a mathematics examination one teacher may give the pupil credit if he uses the correct method in solving a problem, whereas another may give him no credit unless he also carries through the calculation correctly. It has been shown, for example, that the marks given by teachers on papers in English composition are strongly affected by the quality of penmanship in which the papers are written. Whether this is a conscious or an unconscious influence, it doubtless differs with different teachers.

In the second place, the mark which the teacher assigns will depend upon his evaluation of a particular grade of achievement. This variation in the evaluation is responsible for the familiar fact that some teachers are high markers and some are low markers. The same mark of achievement means a different grade of achievement in the minds of different teachers.

A standardized test gives more uniform measures of achievement because, in the first place, it separates these two factors. It does not at the same time attempt to assign a numerical measure to the pupil's achievement and use a mark which expresses an evaluation of this achievement. It gives the pupil's achievement a numerical score and leaves the evaluation to a separate procedure.

The second cause of the greater uniformity and accuracy in standard scores is that the scoring of the pupil's response is systematized and reduced to definite rules. In many cases these rules are so definite and specific that the pupil's paper can be scored in a purely mechanical way by reference to a scoring sheet. In such cases the scorer does not need to pass judgment upon the performance of the pupil but needs only to follow the rules of procedure which are laid down. This routine method of scoring depends upon the organization of the material in the test into certain types of questions. The invention of these types of questions constitutes one of the chief steps in the development of standard tests. These forms of questions will be illustrated in a later section. In other cases the form of the test is not so mechanical that the scoring can be done in this routine fashion, but rules for scoring are laid down which reduce much of the variability between different scorers.

Norms for Comparison. It was said in the preceding paragraph that the variation in marking by means of ordinary examinations was due to the fact that the mark included an evaluation of the pupil's performance as well as an attempt to measure the performance. In the case of standard tests the process of evaluation is separated from that of measurement. The evaluation is based, for the most part, upon the comparison of a pupil's performance as represented in his score with the performance of pupils in general of his same age or grade. This comparison is made possible through the establishment of norms of performance. The norm is simply the average performance of a given group of pupils. It may be the average of all pupils of the age or grade to which the individual belongs, or it may be the average of a more restricted group.

The word norm is not used in the sense of a standard to which the individual should conform. There are some conditions under which we should not expect the pupil to do as well as the general norm, and there are other conditions under which we should expect him to do far better. Furthermore, the norm of present performance does not necessarily set the standard for the future performance of pupils in general. With improved methods or under improved incitement it may be possible to raise the norm beyond its present level. On the other hand, it is also possible that the present norm is higher than is warranted by the practical value of the ability which is represented by it, in the light of the efforts required to attain it. The norm, then, must be evaluated in the light of many conditions. But notwithstanding this fact, the norm constitutes the fundamental basis for the evaluation of the performance of the individual pupil.

The Measurement of Reliability. In the early mental tests, and in school examinations alike, the question of reliability was not raised. In the case of mental tests it was assumed that the score which an individual made on a test was a fair measure of his attainment in the ability which the test measured. This assumption was made plausible by the character of the early tests. When one measures the ability of a person to distinguish between the pitch of two tones, or when one measures the rapidity with which a person can react by lifting his finger at a signal, or when one measures the accuracy with which a person can trace a line, it seems at first glance rather obvious that the score is an accurate measure of a person's ability within the limits of the units of measurement which are used. The assumption that marks on an examination possess any such accuracy can hardly be said to possess this plausibility. In fact, students have been known for generations to express grave doubt concerning the accuracy of school marks. The question was not given serious consideration by educational authorities in part because no convenient technique was at hand for determining the reliability of the marks.

The advent of the correlation technique has supplied the necessary method for determining the reliability alike of mental tests, of examinations, and of educational tests. The use of the correlation method has also given to the term reliability a special and precise significance. By reliability is now meant the consistency of a test. It signifies the degree of agreement between the scores of the same individuals on the same tests repeated or on two forms of the same test. The reliability of a test measures the extent to which the administration and scoring of the test have been so well standardized that the same individuals will attain the same scores when the test is repeated.

Objectivity of scoring, which has already been discussed, is one factor in the reliability of tests. It is, however, by no means the only factor. Even tests in which the score is highly objective may give variable results from time to time, due to the lack of complete standardization of the procedure which the individual follows in taking the test. Large variations may appear in the reaction of a subject from time to time in even such a simple process as tracing a line. Such variations may be produced by changes in the attitude and attention of the subject, by changes in the height of the board on which he is tracing, by changes in the position of his hand and arm, and by variations in the speed of movement. By regulation of such factors as these, the reliability coefficient of the tracing test in an unpublished experiment was raised from about .30 to about .65.

The source of the unreliability of scores on tests may be briefly enumerated. In the first place, the test may yield differing scores on successive administrations because of variations in the general conditions of the test and in the mental set of the individuals who are being tested. The standardization of a test involves a careful study of all such conditions and the development of a set of instructions which will, so far as possible, induce a constant attitude of mind.

A second source of variation, of course, is a lack of uniformity in the standards of scoring or in the application of these standards. This source of error has already been discussed so far as it applies to examinations in general. It is completely eliminated in some forms of standard tests. In other forms it is greatly reduced but cannot be avoided altogether. In the so-called "product" scale, as in the case of the handwriting scale, it is impossible to develop absolute uniformity in the scoring of papers.

A third source of error or unreliability in a test consists of those variations in the behavior of the individual which occur from time to time, other than those which are induced by a change in the mental set. Experiments indicate that variations occur in even the simplest reflexes. Specific causes of these variations have not been determined, but it is a reasonable hypothesis that they are due to the great complexity of the human organism and to a shifting in its internal adjustments and relationships. These constitute a factor in variability which cannot be controlled.

A fourth source of variation consists of sporadic errors. Such errors are due to arithmetical blunders in scoring, unnecessary misunderstanding of directions by the pupils, time lost due to broken pencils, and so on. Errors of this sort may be eliminated by care in the administration of the tests.

These four sources of errors will occur even when the identical test is repeated. When a second form of a test is used, there is a possibility of a fifth error. This error is due to the limitation in the sampling of the items which are used in the tests. For example, if exactly the same arithmetic problems were given to pupils the second time, there would be no variation in the content of the test. If, however, a second set of problems is given, even although they are of the same difficulty for pupils in general, they may vary in difficulty for individual pupils, due, for example, to difference in their familiarity with the situation presented in the problems. This source of error, however, is slight in the case of the tests which are otherwise well organized and which are of suitable length.

All of these causes of variability, and any others which may exist, are measured by the coefficient of correlation between the scores of a group of pupils on one administration of a test and on a second administration of the test, whether the same or different forms are used. The test, of course, should be as reliable as possible. If the coefficient of reliability is not above .80, the test may be considered relatively unreliable. In the case of some of the better standardized tests the coefficient of reliability is above .90.

The Validity of Tests. The determination of the reliability of a test concerns only the question whether the same ability is being measured in successive tests or in successive administrations of the test. It does not touch the question of what ability is being measured. This is the question which is raised when we inquire into the validity of a test. One aspect of this question has already been touched on in discussing the limitations of school examinations as contrasted with standard tests. We found that an ordinary examination in the subject may be very limited because it may cover only a part of the topics which might appropriately be included, and because it may even include some parts which are not important. The standard test corrects this defect by making a careful investigation of the subject in order to determine all those aspects which ought to be included in the test. When the collection of materials for a test is based upon such a careful study, it may be said to be more valid than is a test or an examination the materials of which have been collected in more haphazard fashion. The test is more valid because it is a more adequate measure of what it pretends to measure.

The foregoing statement does not provide us with any definite or mathematical procedure by means of which we may measure the validity of a test. It gives us an inferential basis for judging roughly whether a test is valid but does not give us a means for determining mathematically the degree of its validity.

The method of correlation provides us with an instrument which gives us the theoretical solution of our problem. Just as we can measure the consistency of a test by finding the correlation between successive scores of the same pupils on the test, so we can determine its validity by finding the correlation between the scores on the tests in question and some other and true measures of the ability which the test is supposed to measure. Thus, if we have a true measure of mechanical ability we can correlate the scores on a new mechanical ability test with this true measure, and the size of the correlation coefficient will measure the extent to which the scores on the new test correspond with true measures of mechanical ability.

This method of determining validity is quite satisfactory so far as the mathematical theory is concerned. As soon as we attempt to apply it, however, we immediately encounter difficulty. The fundamental difficulty is the obvious fact that there is usually not at hand a satisfactory measure of the ability which the test is designed to measure. The purpose of the test is ordinarily to supply the lack of such a measure. If we already had it, we would not need to design a test. In some cases, to be sure, a reasonably satisfactory and generally accepted measure may be available, but it may be desirable to construct a more convenient or a cheaper test. In such cases the existing test may be used as a criterion with which to compare the new one. This procedure has often been followed in constructing a group intelligence test. The Stanford revision of the Binet Scale has been taken as a criterion with which the group tests have been correlated.

When an accepted test does not exist in the field of a new test which is being standardized, we may follow one of two courses. In the first place, we may rest content with measuring the pupil's attainment and letting the attainment which is being measured be defined by the nature of the test itself. This is the procedure which is commonly followed in the case of the ordinary subject-matter tests which are used in the schools. In such cases as tests of arithmetic, of reading, of handwriting, of spelling, of history, of foreign languages, and so on, we do not attempt to measure abilities in the general sense of the term but rest content with measuring attainment of a specified sort.

The case is different, however, when we attempt to measure such abilities as musical ability, mechanical ability, quickness or accuracy of movement, or when we attempt to predict the rapidity with which a pupil will master a given subject or a given type of learning before he has begun to learn it.

One method which is sometimes employed in validating a test is to correlate the scores in the test with the teacher's estimate of the abilities of the pupils or with school marks. The logic of this procedure may seem questionable, since the test is supposed to be a more accurate measure of the ability in question than are estimates or marks. The case is not so bad, however, as it seems at first sight. It is quite reasonable to assume that the closer a test comes to an imperfect measure of the ability, the better the test is, and at the same time to assume that the test is a more accurate measure than the criterion to which it is compared. In order to take this position we need only to assume that the errors in the school marks or teacher's estimates, or whatever else is used in the criterion, are as likely to be in one direction as in the other, and that such errors are fewer and smaller in the new test than in the criterion. On this assumption we can conclude that the nearer we can bring a standard test to correlate with even an inaccurate criterion the more valid the test will be.

The other procedure is to give a test at the time of beginning some form of learning and then correlating its scores with the rate of progress which is made. This procedure may be followed, for example, in the case of tests of typewriting ability or tests of musical ability. A test given at the beginning of a course of learning is called a prognostic test. Such a test was given during the war to prognosticate the rate of progress of students in flying. The rate of progress consisted in the number of days necessary to learn to fly independently. A somewhat similar measure with which the score on the test may be correlated is the production record of the individual. This criterion is sometimes used as a measure of the validity of tests of ability to perform factory operations.

While it is not always easy, then, to measure validity in a clean-cut fashion, validity may be determined by indirect methods, and the concept of validity is an important contribution of standardized tests.

Detailed Characteristics and Methods of Standardization

Types of Standard Tests. Standard tests may be classified according to their form into three types. These types may be designated briefly as unit-task tests, short-answer tests, and product scales. Each of these types will be described and illustrated.

Unit-task tests. The unit-task tests, as the name implies, consist of examples of the tasks which are required of pupils in a particular subject. These tasks appear in the tests in the same form as in ordinary instruction. They are selected because of their representative character, and they may be graded according to their difficulty. They may also be grouped into classes according to their specific form or nature.

A clear example of the unit-task test is the spelling scale. The spelling of each word constitutes a unit task. These unit tasks are ordinarily graded according to their difficulty and classified on this basis, as in the Ayres Spelling Scale. In this scale the words are grouped in columns, the words of each column being of approximately equal difficulty. At the top of the sheet on which the test is printed is given for each column the percentage of the words which are correctly spelled in each school grade.

A second example of the unit-task test is the arithmetic test. Both the test in the fundamental operations — adding, subtracting, multiplying, and dividing — and the test consisting of reasoning problems are of this type. There is some difference in principle involved in the tests in fundamentals and the reasoning tests, however. There is a limited number of examples in the fundamental operations if we confine ourselves to the possible combinations up to a certain sized number. There is, however, an unlimited number of possible reasoning problems in arithmetic, since the same number of operations may be performed in dealing with a great variety of concrete situations. When we measure the ability of a pupil to perform a certain combination, then, we are measuring something more precise and definite than when we measure the ability to solve a given reasoning problem.

The standardized reading passages of an oral reading or a silent reading test constitute unit tasks analogous to the reasoning problems in an arithmetic test. We may, therefore, call this kind of reading test a unit-task test. Other examples of this type of test are the language-error test, in which the pupil is required to mark or to correct errors in language usage, and tests in algebra and geometry consisting of typical operations which the pupil is required to perform.

Short-answer tests. The second type of test which is to be distinguished on account of its form is the short-answer test. This may be used to test the ability of the pupil to give information or to pass judgments. In reality it is a collection of examination questions put in such fashion that the pupil is required to give a very definite answer on a very particular point. It is particularly useful in the so-called content subjects, such as history, geography, and science.

There are three principal forms of these short-answer tests. The first is the completion test, sometimes called the recall test. It may be illustrated by the following question in American history:

The name of the man who discovered America was ————————.

The second form of queston is the multiple-choice type. A sentence is given with several possible endings and the student is to designate which ending is the correct one. From three to five may be used. The following is an example:

The telephone was invented by (Edison, Bell, Westinghouse, Maxim).

The third type is similar to the multiple choice except that only two choices are offered. This may be called the alternative-choice type of question. A variant of this type is the true-false question. In this case the complete statement is given, and the student is required to designate the statement as being true or false. The question is usually written in this form:

Washington crossed the Delaware in 1790. TRUE–FALSE

A fourth type of question, which is not so common as these three chief ones, is the matching question. For example, a series of events may be given in order and a series of dates given, including those which belong to the series of events. The series of dates may contain the same number as the series of events, or it may contain a larger number. In any case they are not in the same order as the events. The pupil is required to designate which date goes with each event.

In addition to their use in testing information and judgment in the content subjects, these short-answer tests have been used to measure ability in the use of language. For example, the multiple-choice test may be used to test the pupils knowledge of the meaning of words. For this purpose each word on which the pupil is to be tested is printed with a number of words following it, and the pupil is to designate the one of the following words which has the same meaning as the first word. The true-false type of test may be used by presenting a series of statements and asking the pupil to designate them as true or false. To make this a test of language the subject-matter of the statement should be familiar to the pupil. These methods may be used to test knowledge of a foreign language as well as of a vernacular. A test of translation from a foreign language may be made by giving a sentence in the vernacular and then the translation in the foreign language with one word omitted. The pupil is to supply the omitted word.

The completion and the multiple-choice type of tests are more reliable than a true-false or alternative test of the same number of items. It is probable also that the first two types mentioned are more valid than is the alternative test.

Product scales. The third general type of standard test is the product scale. This test is designed to assist the examiner in judging the quality of the product of the pupil's performance. It consists of a series of specimens of the product of regularly increasing merit. By comparing the pupil's product with the specimens of such a scale, it can be given a score which represents a known degree of merit within the limits of error of judgment. This type of scale is sometimes called a quality scale. Examples of it are to be found in the various handwriting scales, in the English composition scales, and in several drawing scales. The procedure by which the specimens in such a scale are standardized and the steps in the scale determined will be presented in the next section.

General and analytical tests. Besides distinguishing tests according to the form in which the elements of the test are presented, we may distinguish between general tests and analytical tests. General tests are those which aim to measure several forms of achievement together without distinguishing between them, while analytical tests are those which are designed to measure narrower forms of achievement. The most general type of educational test is the general achievement test. This test contains parts which aim to measure the achievement of the pupil in the several school subjects, but it provides a method by which the achievement in all the subjects may be combined to give a single measure.

On the other hand, we may test the achievement in each subject separately. That is, we may test spelling, reading, arithmetic, language, history, and so on, each for itself. If we wish to make a still more analytical examination, we may test the various forms of achievement within a single subject. For example, in the case of reading we may test word meaning, sentence meaning, and paragraph meaning. In the case of arithmetic we may test each of the fundamental operations as carried out with pairs of numbers. We may then test longer operations such as column addition or the subtraction of numbers of more than one digit. We may test addition or subtraction of fractions, long division, the ability to solve reasoning problems, and so on. In the case of handwriting we may test general merit or general appearance, or we may distinguish the quality of writing into its various elements.

A scale which analyzes the achievement of the pupil in the various aspects of a subject is sometimes called a diagnostic test. The purpose of such a test, as the name indicates, is not merely to indicate how well a pupil does in the subjects as a whole, but also to indicate his strong and weak points. Such a diagnostic test is of special value in remedial instruction. It enables the teacher and the pupil to determine the particular weaknesses of the pupil so that practice may be applied at the point where it is most needed.

Scaling Tests. The first form of scaling which may be mentioned is the arrangement of the items of a test in the order of their difficulty and the designation of their relative difficulty by some form of numerical quantity. The purpose of such scaling may be merely to secure items which are equally spaced in difficulty or it may also be to weight the pupil's performance in terms of the difficulty of the items which he passes.

The basis for determining the relative difficulty of the items of a test is the percentage of pupils who pass each of the items. In order to select and arrange the items in roughly equal steps of difficulty it is sufficient to provide that the differences in the percentage of individuals passing successive items are equal. If, however, we assume that the degrees of ability follow the normal distribution frequency, and, if we wish to make the differences between successive items more exactly equal, we may follow a more elaborate procedure. This procedure assumes that equal differences in ability are represented by equal distances along the base line of a probability curve rather than to equal differences in the percentage of failing. The percentage of pupils failing, corresponding to various distances along such a base line, may be found in such a table as the one given by Rugg (49, p. 96).

The second type of scaling is used in constructing a product test. Some form of reasonably accurate scaling is essential in the standardization of this type of test.

There are two general methods of determining the quality of the units of a product scale. The first is the method of equally often noticed differences. This method was used by Thorndike in the construction of his handwriting test. The merit of the handwriting specimens which appear in Thorndike's Scale is based upon the judgment of a considerable number of judges. The difference in quality between one pair of specimens was assumed to be the same as the differences in quality in another pair when the same percentage of persons considered one of the first pair to be superior as considered one of the second pair to be superior.

A different method was used by Ayres in scaling the specimens of his handwriting scale. Instead of basing the merit of the specimens upon the judgment of competent judges, he based it upon the measurement of the legibility of the specimens. After measuring the legibility of a large number of specimens in terms of the amount of time required to read them, he selected specimens which should be equally spaced on a scale of legibility. In determining the spacing between the specimens, he applied the same principle which was described in connection with the method of determining the difficulty of items of the test. That is, he considered equal differences in legibility to be represented by equal distances along the base line of the normal distribution frequency of scores in legibility.

Problems in Scoring. The first problem to be met in scoring tests is to determine the aspects of performance on which the score is to be based. In other words, we must answer the question, what aspects of a performance make it better than another performance?

Speed as a basis of scoring. Obviously one characteristic which leads us to rate one performance as superior to another is speed. Speed is of greater importance in some kinds of performance than in others. Speed is of relatively great importance in more routine types of performance than it is in those which involve originality and invention. Thus, speed is of considerable importance in handwriting, in arithmetical calculation, in typewriting, in simple factory operations, in laying bricks, or in washing dishes. Speed is of relatively less importance in such activities as inventing a new machine, discovering a new scientific principle, writing a book, or diagnosing a disease. In these cases the quality of the performance overshadows in importance its speed.

In those cases in which speed of performance is of major importance, the score may emphasize speed as one of its chief factors. This may be done by using one of two measures. In the first place we may measure the amount of time required to perform a given amount of work, and, in the second place, we may measure the amount of work performed in a given unit of time. While these two measures both represent the speed of performance, they are not identical. The one is the reciprocal of the other. Furthermore, they do not give the same learning curve, or the same curve when the scores are plotted by ages or by grades. It is not possible to say that one of these measures is right and the other wrong, but it is necessary in comparing different curves based on speed to be sure that the speed is reckoned in the same fashion.

While neither the time measure nor the amount measure may be considered right as distinguished from the other, each has certain practical advantages. To have all individuals who are being tested do the same amount of work and then to measure the time of performance insures that all the individuals shall have attempted exactly the same problems. Hence, for example, variations in difficulty will not be so serious as they would be if the different individuals covered different amounts of ground. In order to secure this measure, however, it is necessary that each individual be timed separately. This is an awkward thing to do in a group test. It is much easier to give all the individuals the same amount of time and then to record the score in terms of the number of items which each individual has passed. This gives us the amount score. If the items on the test are of equal difficulty or of well-graded difficulty, the amount score is a convenient and reasonably satisfactory one to use.

Accuracy in relation to speed. It will doubtless have occurred to the reader that it is necessary to take account of more than speed in measuring performance in the above examples. One would hardly judge the performance of an individual in adding columns of figures solely by the number of columns which were added per hour. The person who has added 20 columns and got them all right has made far better performance than a person who has added 20 columns and got three-quarters of them right. He has made a far better performance even than the person who has added 30 columns and got 20 of them right. In the case of arithmetical calculation, accuracy is of prime importance along with speed. In fact, if the two were not correlated, as they are in fact, accuracy would be of considerably greater importance than speed.

One of the problems in scoring tests is to determine how to take account of accuracy and speed. One method is to credit the individual with only the items which he attempted and got correct. The score would then consist of the number right rather than the number attempted. This may be a satisfactory method of scoring in those cases in which accuracy is of minor importance. In the case of arithmetic, however, and in the case of a good many other operations, this does not give sufficient penalty for errors. Another method might be to require the individual to check his answers until he was sure they were all correct and then measure the time required. This is a method which is frequently used in clerical occupations, but it has not been used to any great extent in educational tests. A third method is to base the score on the number right with certain deductions for errors. In measuring typewriting efficiency, for example, there are deductions which are conventionally established for various types of errors. There has been little scientific study made of the amount of deduction which should be made for errors. But Thurstone (61) has devised a procedure by means of which the allowance can be calculated.

Quality of performance. The second general aspect of performance upon which the score may be based is its quality. Quality is usually measured by means of a scale of merit. It is illustrated by quality in handwriting, as measured by either the Thorndike or the Ayres Scale, by quality in written composition, or by quality in drawing. As the illustrations indicate, quality is usually measured by means of a product scale. It might theoretically be possible to analyze quality into its elements and to measure each one of the elements, but this would be a very complicated and difficult thing to do.

Difficulty of the task. A third general aspect on which scores are based is the rank of the performance on a scale of difficulty. We have no one very satisfactory word to designate this aspect of performance. Thorndike (58, chap. 1) called it the altitude of performance. It has sometimes been called, rather less aptly, the power of performance. Tests which measure differences in this aspect of performance have been called power tests. This aspect is measured by means of a scale consisting of items of regularly graded difficulty. If it were the only basis of judgment, the individual's score would consist of the number of items along such a scale which he was able to pass. Ordinarily, however, there is a time limit, so that speed is also a minor factor in the score. A test which is organized to measure altitude or power is valuable in measuring complex types of performance.

Types of Scores

Raw scores. The individual's performance on an educational test is represented by a number. In the case of a product scale this number is the one which is attached to the step on the scale matching the performance. In a unit-task scale it may represent the number of items passed in a given time interval. This number is called the raw score. The raw score, as has already been said, means nothing in itself. One must interpret it by comparing it with a norm. A raw score of 50 may be a high score if the norm, or the average score of children of the same age as the one making the score, is 40. If, on the other hand, the norm is 60, the score 50 is a low one.

Percentile scores. A more accurate method of interpreting the individual's score is to relate it not simply to the norm but to the entire distribution of scores. One method of doing this is to find the percentile score. The percentile score is that score below which a certain percentage of the individual scores fall. For example, a percentile score of 30 is a score below which 30 per cent of the scores are found. A percentile score of 50 is, of course, the median score. A percentile score has the advantage of meaning the same thing whatever the particular units of measurement are and whatever the test which is used. It has the disadvantage of assuming a rectangular distribution instead of a normal distribution of scores. This, however, is not a serious disadvantage in view of the somewhat crude scores which are derived from most educational tests.

Standard scores. A second method, which is statistically more accurate, is to express the individual's score as a fraction of the standard deviation of the distribution of all the scores, expressed as the positive number for the scores above the average and a negative number for the scores below. This is a measure which has the same meaning for all tests and for all units of measurement. It is the most accurate measure known. Scores on different tests expressed in this unit can be combined by simple addition and averaged.

Age and grade scores. The methods which have been described are for the purpose of expressing the individual's score in relation to the distribution of scores of his own group as, for example, his age group or his grade group. It may be desired to express the score in terms of a more general scale. One method of doing this is to make a scale of the average scores of pupils of different ages and then compare the individual's score with this scale. Such a scale is analogous to the mental-age scale for scoring intelligence tests and may be called an educational age scale. The pupil's score, then, may be expressed as a subject age if it is a score in tests of a single subject, or as an educational age if it is a score in all subjects taken together. A variation of this method is to use grade groups rather than age groups. The score expressed in terms of the school grade to which it corresponds is sometimes called the B score. The value of grade scores is less stable than that of age scores, however, because it is affected by differences in rate of progress of children through the school.

The "T" score. Another method of scoring, which gives a scale with common units, is the so-called T score described by McCall (35, chap. X). This method is based upon the distribution of scores of children of the 12-year-old age group. The unit of measurement and of scoring is one-tenth of the standard deviation of the distribution of scores of this group. For convenience in scoring and to avoid negative scores the average score of the 12-year-old group is set arbitrarily at 50.

Quotients as scores. There has been some use in educational tests of a quotient which is analogous to the intelligence quotient used with intelligence tests. This quotient expresses the relative attainment of the individual in comparison to others of his own age group. It is calculated by first finding the individual's score in terms of age norms and then finding the ratio between this score, called the Educational Age or E.A., and the individual's chronological age. If the test score measures achievement in a single subject, this ratio is called the Subject Quotient. If it measures achievement in a number of subjects, the ratio is called an Educational Quotient, or E.Q. for short.

The expression of the child's intelligence in terms of the I.Q. and the expression of his achievement in terms of the E.Q. has suggested that it might be possible to express the relationship between achievement and ability by the ratio of the Educational Quotient to the Intelligence Quotient. This is expressed either as $\frac{\text{E.Q.}}{\text{I.Q.}} = \text{A.Q.}$ or $\frac{\text{E.A.}}{\text{M.A.}} = \text{A.Q.}$ A.Q. in these equations stands for Achievement Quotient. The Achievement Quotient has variously been taken to represent the relation of achievement to ability, the relation of a child's actual achievement to his possible achievement, or the relationship of a child's achievement to the achievement of the average of children of the same ability. It has also sometimes been taken to be a measure of effort. The fact that we frequently get achievement quotients which are above 1.00 indicates the absurdity of taking it to represent the relation between achievement and capacity to achieve, since it is obvious that nobody could achieve beyond his capacity.

The value of the Achievement Quotient as a measure of any well-defined thing is coming to be doubted on two grounds. In the first place, there is not the clear-cut distinction between the measurement of achievement and the measurement of capacity which is often implied. In the second place, the interpretation of the Achievement Quotient at its face value implies that there is a perfect correlation between capacity to achieve in general and capacity to achieve in those particular respects which are covered by the educational tests. This perfect correlation certainly does not exist in respect to some of the individual school subjects and probably does not exist in respect to all of them taken together. In any case the meaning of the Achievement Quotient is not at all clear, and if it is used at all it must be used only as a very rough approximate indication.

Correction for "guessing." In one of the particular forms of test question which has been mentioned, namely, the true-false or the alternative type of question, a special procedure in scoring has been used. This procedure is to make allowance for the possibility of getting the correct answer in half the cases by mere guessing or random response. This method involves taking as the score not the number of right responses but the number of right responses minus the number of wrong responses. The scoring formula becomes $R-W=S$. The assumption underlying this scoring formula is that if a person makes random responses on a certain number of items he will get half of them right and half of them wrong. The errors, therefore, are to be considered as the wrong half of these random responses, and an unequal number of right responses are to be considered as random responses or as due to guessing. When this method of scoring is used, the persons taking the tests are frequently instructed, "When in doubt, guess." Guessing is thus encouraged. There is some question on both psychological and statistical grounds of the propriety of using this scoring method. It is simpler, psychologically sounder, and possibly statistically as accurate to discourage guessing and then take as the score the number of right responses merely.

Weighted scores. At one period in the testing movement a good deal of attention was given to weighting the test scores. Weighting is of two sorts. When a scale or test is a composite of a number of tests, each one of which measures a different subject or a different phase of the subject, the scores on the different parts are sometimes weighted in making up the composite score. This is usually done in the cases in which the different parts of the composite tests have varying numbers of items so that the possible scores on the different parts vary. The weighting in this case is made in order to equalize the share of the parts in the composite. In another type of weighting the individual detailed items of a test are weighted in accordance with the difficulty of these individual items. The principle underlying this type of weighting is that the more difficult items ought to yield a greater share of the total score than the less difficult items. The first type of weighting is rather easily carried out and does not involve much additional labor. While it is not certain that it adds to the accuracy of the composite score, it may be worth while. It is questionable, however, whether the second type of weighting adds enough to the accuracy of the score to compensate for the large amount of labor which is entailed.

Norms. It has already been pointed out that the use of norms is characteristic of standard tests and is the essential means by which the scores of the individuals are interpreted. Mention has been made of the fact also that norms are sometimes established by grade groups and sometimes by age groups. We may add a brief comment on these two types of norms.

Age norms are the ones most commonly used in interpreting the scores of intelligence tests. Age norms are established by tabulating together the scores of children of a given age and finding the average of these scores. In this way averages are obtained for the children of each successive age. Various methods may then be used, as has already been described, of comparing the score of the individual child with the norm for his age. The use of age norms in intelligence tests is due to the assumption that intelligence tests measure natural ability and the growth in natural ability which comes from increasing age rather than from education. Whether a child is in one school grade or another is not supposed to make much difference in his intelligence, but his intelligence is determined solely by a combination of his mental capacity in comparison with the average and his degree of maturity as indicated by his age.

In the case of educational tests, on the other hand, it has been considered more appropriate to use grade norms. This is for a similar reason to that which leads to the adoption of age norms in intelligence tests. It is assumed that the score in achievement tests is due largely to the training which the child receives and that the amount of training he has had should, therefore, be the basis for the norms with which the child's score is compared. For this reason the scores are classified by grades and the child's score is compared with the norms in the school grade in which he is situated.

It is probable that neither of these assumptions is entirely justified. The child's score on an intelligence test is determined partly by the amount of training he has received. On the other hand, his score on the educational test is determined in large measure by his intelligence. The use of both types of norms, therefore, involves a certain amount of error. For practical purposes, however, they have been found to be serviceable. We will avoid undue error by refraining from drawing too far-reaching theoretical conclusions from the measures which express the relationship of the child's score to the norm.

For some purposes it is desirable to use general norms and for some purposes local norms or norms for special groups. General norms are those whcih are based upon the tabulation of scores of children from all localities and all classes of community or of schools. These norms may be called country-wide norms. Local norms are norms based upon the attainment of children in a given city or town or even in a given school. One may even use class norms. General or country-wide norms are of greatest value in educational tests when we wish to compare the achievement of the children of a community as a whole with the achievement of other communities. Such a comparison may be made to reveal factors which affect the achievement of the children in a given community and which differentiate between that community and others. When we are interpreting the achievement of an individual child, however, it may be misleading to compare his score with the general norm, if this norm differs widely from the norm of achievement of the local group to which he belongs. The factors which affect the entire community in which a child lives will, in some measure, affect each individual child. What we are seeking in scoring the individual child is not to get at these factors which are common to the group but rather to get at the factors which differentiate this individual child from the other children of the group to which he belongs. Local norms, therefore, are useful in diagnosing the individual child, whereas general norms are useful in diagnosing or interpreting the achievement of the children of a community or of a school taken as a whole.

The Uses of Educational Tests

Surveys. While the purpose of a chapter on educational tests in a general book on psychology is not to emphasize the practical application of tests, a brief summary may be made of the practical uses to which tests have been put. The first important use of educational tests is illustrated by the work of Rice, who is the pioneer in this field. Rice devised his spelling test for the purpose of comparing different cities with each other. This comparison is called an educational survey. It is obvious that we need some such common basis of comparison as is furnished in an educational test to enable us to compare the achievement of pupils in different communities with each other. A large number of surveys of cities, of counties, and even of whole states have been made. These surveys have utilized educational tests and have at the same time stimulated the development of these tests. Modern educational systems use the survey not as an occasional and sporadic attempt to check up upon the efficiency of the school system but rather as a means of making a continuous study of the efficiency of the system and of the means of improving it. For this purpose practically all school systems of sufficient size have a department of educational measurements manned by experts in this field.

Research. A second important use of educational tests goes beyond the mere checking up of the efficiency of a school system. It serves as a tool for research. An obvious type of research is the comparison of different methods of instruction. The effectiveness of different methods of instruction can be measured most directly by measuring the achievement of the children who are taught by the various methods. If an educational test gives objective and comparable measures of what the children have learned, it constitutes the best instrument for the comparison of various methods. Accordingly, educational tests have been used extensively in large-scale educational investigation. A recent example is the modern foreign language study which was organized under the auspices of the Carnegie Corporation and the sponsorship of the American Council on Education.

Evaluation of Teaching. Some attempts have been made to use educational tests as means of comparing the efficiency of different teachers. This procedure rests on the assumption that the efficiency of the teacher may be measured by the attainments of the pupils. This assumption is probably correct in the long run, but the measurement is complicated by the fact that the attainments of a given group of pupils are affected by a variety of elements or factors. If one teacher had a group of pupils throughout their entire school career, this might be a satisfactory method. Due to the complication of conditions, however, it has not proved to be a very useful one.

Motivation. A more important use is the objectification of teaching and the motivation of learning. It is a well-established principle in learning that the individual progresses more rapidly when he knows definitely what he is aiming at and when he is able to measure his progress in exact terms. Educational tests provide the means for making this possible in the school. They enable the pupil to recognize his attainments precisely in comparison with the attainments of other pupils and with standard attainments. If this information is tactfully used, it proves a powerful stimulus to the child. It is found useful not only to compare the child at a given moment with the norm but also to make a series of tests and to reveal his progress to him. Whether, then, he stands at the top or bottom of the class, he is able to see the effect of his effort by a comparison of his attainment from time to time. The use of educational tests, then, has greatly influenced the methods of instruction in certain of the subjects, particularly those which involve the development of skill and routine practice.

Individual Diagnosis. Implied in what has been said is the use of educational tests for individual diagnosis. They are used for this purpose in conjunction with intelligence tests. While the use of the Achievement Quotient is attended with the limitations and dangers which have been mentioned, it is found useful to make a general, rough comparison between the pupil's score upon a test of general ability and his score upon a test of his educational attainment. Pupils are frequently discovered by such comparison who are apparently capable of doing better work in the school than their attainments indicate. When such pupils are discovered by means of tests, it is sometimes possible to unearth and to remove the causes of their low attainment. The cause may consist in poor social adjustment as represented, for example, in discouragement or in an attitude of sullenness and of obstinacy. In some cases such diagnosis reveals specific defects in the child's previous preparation. A little special practice frequently serves to remove these defects and to improve greatly the child's present attainment.

The Constitution of Special Abilities as Revealed by Tests

We come in this final section of the chapter to a topic which is of the most importance, of any which are here discussed, for general psychology. Unfortunately, however, our information on the constitution of special abilities is so uncertain and the identification of special abilities is fraught with so many statistical difficulties and hazards that we can do little beyond making tentative suggestions concerning this important matter.

A procedure which has often been followed in attempting to measure and classify special abilities is to sidestep the main difficulty by taking the score on a particular test as a measure of a distinct ability. This, of course, would mean that there are as many abilities as there are tests, and it would make the diagnosis of abilities depend upon the fortuitous construction of tests.* This obviously is not a sound procedure. Certain general facts regarding the distribution of abilities may, however, be gleaned from the results of current tests.

Distribution of Abilities. There are certain facts which may be stated concerning the scores of groups of individuals on educational tests. For example, we may plot the distribution of these scores. When this is done we find that the distribution commonly follows the normal frequency curve. This is characteristic of tests of all sorts, when the population may reasonably be supposed to be a random sample, and when the test is of such difficulty as to give opportunity to individuals at both ends of the scale to make a score which comports with their ability. From this fact of the normal distribution of scores we may perhaps infer that the special abilities, if they could be measured, would also follow the same distribution. If, in similar fashion, we take the scores on educational tests as measures of special ability, without raising for the moment the question whether they measure special abilities or composites of abilities, we may consider certain other relationships concerning the measures of these abilities.

A second question which has been raised concerning the distribution of abilities, in addition to the form of the distribution, concerns its range. Attempts have been made to determine how much better the best individual is than the poorest in the ability measured by a given test. This has been done by finding the ratio between the lowest score and the highest score. By this method it has been found that the highest score may be from one and a half to twenty or more times as large as the lowest score. It would be very hazardous to say, however, that the ability of the most able individual in some special ability is one and a half or twenty times as great as the ability of the least able individual. To multiply and divide scores in this fashion implies that we know the zero point of ability, and that a zero score represents this zero point of ability. This, of course, is far from true. Furthermore, the zero point in some scales may have a different relation to the true zero point of ability than does the zero point in other scales. It is, therefore, unwarranted to compare the ratios in the ability of the poorest and the best individual on different scales. Such attempts as have been made are probably worthless.

Correlation and Specialization. In the next place we may find the correlation between scores on educational tests, considering them as special-

TABLE 1

CORRELATIONS OF THE MEAN GROUP TEST AND STANFORD MENTAL AGE WITH ABILITY IN SCHOOL SUBJECTS

Mean of Results for Grades IV to VI

	Reading comprehension	Reading rate	Arithmetic	Spelling	Composite achievement
Mean Group Test*	0.59	0.49	0.27	0.35	0.52
Stanford Mental Age	0.49	0.46	0.30	0.31	0.54

*Data for Myers Test omitted.

TABLE 2

	Arith. comp.	Arith. reas.	Word meaning	Sentence meaning	Paragraph meaning	Language usage	Spelling	Science inform.	Language & Lit.	History & Civics	Music & Art	Inform. total	Reading total
Arithmetic reasoning	.55												
Word meaning	.34	.37											
Sentence meaning	.30	.37	.75										
Paragraph meaning	.25	.36	.54	.47									
Language usage	.13	.10	.45	.45	.23								
Spelling	.54	.45	.54	.46	.46	.35							
Science information	.29	.41	.47	.39	.24	.25	.41						
Language & Literature	.24	.38	.64	.49	.41	.30	.38	.55					
History & Civics	.44	.61	.62	.50	.39	.18	.54	.55	.66				
Music & Art	.30	.36	.59	.47	.34	.40	.44	.51	.60	.54			
Reading total	.36	.43				.44	.57	.43	.60	.60	.54	.63	
Arithmetic total			.44	.42	.37	.14	.63	.43	.38	.64	.41	.88	.46

ability tests, and scores on intelligence tests, considering them as measures of general ability. Such measures do not suffer from the statistical fallacy which is dependent upon the attempt to measure the range of ability. Such correlations, however, are found to be variable, and it is difficult to draw conclusions from them which will be found to be generally applicable. One table (17, p. 279) containing such correlations may be given by way of illustration (Table 1). This table gives some surprising results. It shows, for example, that there is a closer correlation between a test of spelling ability and the scores on intelligence tests than between the score on the arithmetic test and the scores on the intelligence test. One would hardly maintain, however, that ability to spell is a more important constituent of intelligence than the ability to perform arithmetic operations, especially of the type which is characteristic of Grades 4 to 6. These correlations reinforce the conclusion that it is difficult to draw conclusions concerning the existence or amount of special abilities and general abilities from the scores on the educational and the intelligence tests. We may draw conclusions concerning the composites of ability measured by these various tests but the correlations suggest that it is necessary to carry on considerable analysis before we draw conclusions concerning the existence and degree of special abilities.

The facts concerning the intercorrelation of tests of school subjects may be illustrated by Table 2, which is presented by Terman (57, p. 319). We find in this table a wide range of degrees of intercorrelation from the lowest, between language usage and arithmetic reasoning, of .10, to the highest, between arithmetic total and information total, of .88. There are seven correlations of .30 or below and twelve correlations of .60 or above. Casual inspection of the distribution of these low and high correlations does not seem to give any ready explanation of them. If such an explanation exists it must be based upon a thoroughgoing statistical analysis. The methods and results of such an analysis, which aims to reveal hidden factors responsible for the intercorrelation among such tests as these, are treated in another chapter of this book. No further attempts at analysis will be made at this point. We may add, however, some comments concerning the limitations of tests as direct measures of special ability, and finally a few comments concerning the probable existence of various special abilities.

The Limitations of Tests as Measures of Special Ability. A little acquaintance with tests and a study of their results will convince one that there are serious difficulties in the way of using them as measures of special ability. For example, the subject-matter of a test, whether it be an intelligence test or an educational test, is particularized. If we wish to measure a pupil's knowledge of history, we cannot test him in knowledge of history in general; we must test him in the knowledge of certain particular historical facts. In the same way, we cannot test the pupil in ability in arithmetic reasoning in general, we must test him in the ability to reason out certain particular problems in arithmetic. This constitutes a limitation in all cases except those in which we can test in all the particulars which make up the subject in question, as, for example, in the case of arithmetic combinations.

Another difficulty is that the means which we take to measure an ability may involve other abilities than those which are essential to the operation itself. Suppose, for example, that we test spelling by means of a dictation exercise. The child's score on the spelling test will be affected not only by his knowledge of the spelling of the words themselves but also by his ability to understand correctly the sentence which is spoken by the person giving the test. Again, if we are giving a test in the speed of writing, and the test is conducted by having pupils memorize a passage and then write it from memory, the score of some of the children may be affected by lapses in memory which cause them to pause in their writing to recall what they were to write.

In the third place, a pupil's performance may be affected not only by his abilities, whether special or general, but also by his attitude or feeling toward the circumstances which surround the test. This is a general source of error which tends to lower all correlations. It does not, of course, produce spurious correlations as may some of the factors which have previously been mentioned.

The most general difficulty, of course, in the use of scores on tests as indications of general ability or of special abilities is the complication of factors underlying the performance in the various tests. Any one test, whether designed to measure a special or a general ability, will measure more than the ability which it is designed to measure. A test of special ability will be to some extent a measure of general ability. A test of one special ability will to some extent measure other special abilities. We cannot suppose that the demarkation between special abilities corresponds to the boundary line between tests. We are very much in the dark as to what special abilities exist and what their nature is. It is probable that some tests come nearer to measuring relatively pure abilities than do others. We can make some sort of guess at the existence and nature of a few abilities and must leave to further research the identification of others. In the next section we shall hazard a few such guesses.

Types of Abilities

We may first raise the question whether tests in the various school subjects do measure different abilities, that is, whether the ability measured in an arithmetic test is different from the ability measured in a reading test, or a test in science. A method for determining the percentage of cases in which a difference between two tests represents the real difference in ability, that is, a difference greater than chance, has been worked out by Kelley. Table 3 gives sample sets of such differences as were found by Kelley (31, p. 333). This means, for example, that in Kelley's estimate 34 per cent of the differences between the scores in arithmetic reasoning and the scores in composition may be ascribed to real differences in ability and not to chance. This table means that there is a certain amount of specialization in ability. It does not reveal the nature or constitution of the various special abilities which may exist in the different subjects, nor does it reveal the amount of overlapping among these abilities in the different subjects.

Spearman (53, chap. XIII) takes a sharply critical attitude toward the ordinary notion of the nature and existence of special abilities. He points out, for example, that there is a very low correlation between the scores of tests in various forms of sensory discrimination and that, therefore, we are not warranted in speaking of a special ability in sensory discrimination. To call sensory discrimination a special ability would be, he says, to set up a faculty. He levels the same criticism at the designation of memory as

TABLE 3

Percentage of Differences in Individual Test Scores in Excess of the Chance Percentage in Case the Reliability of Each Test is .824

	Computation	Arithmetic reasoning	Word meaning	Sentence meaning	Paragraph meaning	Language usage	Spelling	Science inform.
Arithmetic reasoning	34							
Word meaning	39	30						
Sentence meaning	40	26	13					
Paragraph meaning	35	26	14	17				
Language meaning	39	33	24	23	17			
Spelling	31	36	28	33	25	30		
Science information	35	20	24	25	21	30	31	
History and literature information	37	26	20	20	20	27	31	9

a special capacity. Again, he points out that Simpson found very little correlation between two tests of the perception of length. The special abilities, according to Spearman, are exceedingly narrow and restricted. These abilities, which he designates *s*, are entirely uncorrelated with each other and entirely uncorrelated with the general factor *g*.

Spearman's criticism of the notion that there exists a special factor of sensory discrimination or a special memory ability may be accepted on the ground of the low correlation between various tests in each of these fields. Whether there do not exist special abilities within the various types of sensory discrimination, for each type of sensation, however, is not so certain. It seems probable, for example, that there does exist a special ability in the discrimination between pitch of sounds, a special ability in discrimination between colors, and so on. This conclusion, however, must be advanced with some reservation.

Attempts have been made to determine whether there exists special motor ability. Here again we find somewhat the same situation as in regard to special sensory discrimination. There is not one special motor ability, but there may be two or three. For example, quickness of movement may represent a special ability; accuracy and steadiness of movement may represent another special ability. There seems to be a rather clear distinction between quickness, on the one hand, and steadiness and accuracy, on the other hand. This conclusion, however, is again tentative.

With respect to memory also there seems to be a division. Tests of rote memory correlate with each other more closely than they do with tests of logical memory. Again, there is a closer correlation between various tests of logical memory and between various tests of rote memory than there is between tests of logical memory, on the one hand, and tests of rote memory, on the other hand. It is doubtless incorrect to speak of memory ability as though it were a single uniform trait. It may be possible, however, to speak justly of types of memory ability.

When we pass beyond these rather elementary types of ability, we arrive at those that are more complex. It is customary, for example, to speak of mechanical ability, and we have several tests of mechanical ability. It is probable that further investigation will reveal that mechanical ability is a composite. It may be possible to identify various special abilities as constituent of mechanical ability. The same thing may be said of musical ability. Seashore has analyzed musical ability into five chief constituents and many other supplementary special abilities. We may for practical everyday convenience speak of musical ability as though it were a simple special ability, but in reality it is a composite.

The same is true of many other special abilities or talents which are commonly distinguished in popular thought. Artistic ability, language ability, ability in science, ability in mathematics, ability in practical affairs, are all made up of an indefinite number of constituent abilities. Of the nature and number of these constituent abilities we know very little indeed. We must await their further analysis and description by future research.

REFERENCES

1. AYRES, L. P. A measuring scale for ability in spelling. New York: Russell Sage Foundation, Division of Education, 1915.
2. ———. A scale for measuring the quality of handwriting of school children. New York: Russell Sage Foundation, Division of Education, 1912.
3. BINET, A., & SIMON, T. Le développement de l'intelligence chez les enfants. *Année psychol.*, 1908, **14**, 1-90.
4. BRONNER, A. F. The psychology of special abilities and disabilities. Boston: Little, Brown, 1917.
5. BUCKNER, C. A. Educational diagnosis of individual pupils. New York: Teach. Coll., 1919.
6. BURT, C. L. Mental and scholastic tests. London: King, 1922.
7. ———. Handbook of tests for use in schools. London: King, 1923.
8. CHADWICK, E. [quoted]. Educational measurements of fifty years ago. *J. Educ. Psychol.*, 1913, **4**, 551.
9. CHAPMAN, J. C. Trade tests; the scientific measurement of trade proficiency. New York: Holt, 1921.
10. CHARTERS, W. W., & MILLER, E. A course of study in grammar based upon the grammatical errors of school children of Kansas City, Missouri. *Univ. Mo. Bull.*, 1915, **16**, No. 2.
11. CORNING, H. M. After testing—what? The practical use of test results in one school system. Chicago: Scott, Foresman, 1926.
12. COURTIS, S. A. Measurement of growth in efficiency in arithmetic. *Elem. School Teach.*, 1909, **10**, 58-74, 177-199; 1910-11, **11**, 171-185, 360-370.
13. FENTON, N. An introduction to educational measurements. Boston: Ginn, 1928.
14. FRANZEN, R. H. The accomplishment ratio. New York: Teach. Coll., 1922.
15. FREEMAN, F. N. An analytical scale for handwriting. *Elem. School J.*, 1915, (Jan.).
16. ———. Mental tests. Boston: Houghton Mifflin, 1926.

17. GATES, A. I. The correlations of achievement in school subjects with intelligence tests and other variables. *J. Educ. Psychol.*, 1922, **13**, 277-285.

18. ————. Manual of directions for the Gates Primary Reading Tests. New York: Teach. Coll., 1926.

19. GILLILAND, A. R., & JORDAN, R. H. Educational measurements and the classroom teacher. New York: Century, 1924.

20. GRAY, W. S. Studies of elementary school reading through standardized tests. *Suppl. Educ. Monog.*, 1917.

21. GREENE, H. A. Work-book in educational measurements. New York: Longmans, Green, 1928.

22. GREENE, H. A., & JORGENSEN, A. N. The use and interpretation of educational tests. New York: Longmans, Green, 1929.

23. GREGORY, C. A. Fundamentals of educational measurement with the elements of statistical method. New York: Appleton, 1922.

24. HEALY, W., & FERNALD, G. M. Tests for practical mental classification. *Psychol. Monog.*, 1911, **13**, No. 2.

25. HENMON, V. A. C. Air service tests for aptitude in flying. *J. Appl. Psychol.*, 1919, **3**, 103-109.

26. HINES, H. C. A guide to educational measurements. Boston: Houghton Mifflin, 1924.

27. HOLLINGWORTH, L. S. Special talents and defects. New York: Macmillan, 1923.

28. HORN, E. The 3009 commonest words used in adult writing. *4th Yrbk. Dept. Superinten. Nat. Educ. Asso.*, 1926, 146-169.

29. HULL, C. L. Aptitude testing. Yonkers: World Book, 1928.

30. JONES, N. F. Concrete examination of the material of English spelling. *Univ. S. D. Bull.*, 1913.

31. KELLEY, T. L. A new method for determining the significance of differences in intelligence and achievement scores. *J. Educ. Psychol.*, 1923, **14**, 321-333.

32. ————. Interpretation of educational measurements. Yonkers: World Book, 1927.

33. ————. Crossroads in the mind of man: a study of differentiable mental abilities. Stanford Univ.: Stanford Univ. Press, 1928.

34. LINCOLN, E. A. Beginnings in educational measurement. Philadelphia: Lippincott, 1924.

35. MCCALL, W. A. How to measure in education. New York: Macmillan, 1922.

36. ————. How to experiment in education. New York: Macmillan, 1923.

37. MONROE, W. S. Educational tests and measurements. Boston: Houghton Mifflin, 1917.

38. ————. An introduction to the theory of educational measurements. Boston: Houghton Mifflin, 1923.

39. ————. Educational tests for use in high schools. (Rev. by C. W. Odell.) *Univ. Ill. Bull.*, 1925, **22**, No. 37.

40. ————. Educational tests for use in elementary schools. (2nd rev. by C. W. Odell.) *Univ. Ill. Bull.*, 1927, **24**, no. 49.

41. MONROE, W. S., DEVOSS, J. C., & KELLY, F. L. Educational tests and measurements. Boston: Houghton Mifflin, 1917.

42. MÜNSTERBERG, H. Psychology and industrial efficiency. Boston: Houghton Mifflin, 1923.

43. ODELL, C. W. Traditional examinations and new-type tests. New York: Century, 1928.

44. ————. A critical study of measures of achievement relative to capacity. *Univ. Ill. Bull.*, 1929, **26**, No. 29.

45. Paulu, E. M. Diagnostic testing and remedial teaching. New York: Heath, 1924.

46. Pressey, S. L. & Pressey, L. C. Introduction to the use of standard tests. Yonkers: World Book, 1922.

47. Rice, J. M. Scientific management in education. New York, 1913.

48. Ruch, G. M., *et al.* Objective examination methods in the social studies. Chicago: Scott, Foresman, 1926.

49. Rugg, H. O. Statistical methods applied to education. Boston: Houghton Mifflin, 1917.

50. Seashore, C. E. The psychology of musical talent. Newark, N. J.: Silver, Burdett, 1919.

51. Smith, H. L., & Wright, W. W. Tests and measurements. Newark, N. J.; Silver, Burdett, 1928.

52. Spearman, C. General intelligence objectively determined and measured. *Amer. J. Psychol.*, 1904, **15**, 201-292.

53. ————. The abilities of man. New York: Macmillan, 1927.

54. Starch, D., & Elliot, E. C. Reliability. of grading work in mathematics. *School Rev.*, **21**, 254-259.

55. Stenquist, J. L. Measurements of mechanical ability. New York: Teach. Coll., 1923.

56. Stone, C. W. Arithmetical abilities and some factors determining them. *Teach. Coll. Contrib. Educ.*, 1908, No. 19.

57. Terman, L. M., *et al.* Genetic studies of genius. Vol. I. Mental and physical traits of a thousand gifted children. Stanford Univ.: Stanford Univ. Press, 1925.

58. Thorndike, E. L. The measurement of intelligence. New York: Teach. Coll. Bur. Publs., 1925.

59. ————. Handwriting. *Teach. Coll. Rec.*, 1910, **2**, (March).

60. Thorndike, E. L., & McCall, W. A. Silent reading scales. New York: Teach. Coll. Bur. Publs.

61. Thurstone, L. L. A scoring method for mental tests. *Psychol. Bull.*, 1919, **16**, 235-240.

62. Trabue, M. R. Measuring results in education. New York: American Book, 1924.

63. Van Wagenen, M. J. Educational diagnosis and the measurement of school achievement. New York: Macmillan, 1926.

64. Wilson, G. M., & Hoke, K. J. How to measure. (Rev. ed.) New York: Macmillan, 1928.

65. Wissler, C. Correlation of mental and physical traits. *Psychol. Monog.*, 1901, **3**.

66. Wood, B. D. Measurement in higher education. Yonkers: World Book, 1923.

67. ————. New York experiments with new-type modern language tests, etc. New York: Macmillan, 1927.

68. Woody, C. Measurement of some achievements in arithmetic. *Teach. Coll. Contrib. Educ.*, No. 80.

69. Woolley, H. T., & Fischer, C. R. Mental and physical measurements of working children. *Psychol. Monog.*, 1914, **18**.

70 17th yearbook of the National Society for the Study of Education. The measurement of educational products. Bloomington, Ill.: Pub. School Publ. Co., 1918.

CHAPTER 19

THE ADULT IN THE COMMUNITY

Mark A. May
Yale University

The "individual" of the three preceding chapters—the hero of this volume—has been studied in infancy and in school and now arrives a full-fledged adult ready to take his place in the community. Much more is known about his early childhood and school days than about his later life. One reason for this is that in infancy and during his earlier years he was more docile, easier controlled, and a much better subject for psychological study than he now is. As we see him in the community he is rather elusive and difficult to manage; he runs away from the laboratory, he refuses to take our tests unless he has first been "sold" on the idea. On the other hand, he does cooperate with us in many ways, with the result that we know some things about him and about how he fits in his community. Samples of these facts will be presented in this chapter.

From Experimental to Individual Psychology. Experimental psychology has been open to the criticism that it is concerned primarily with the study of abstractions rather than with concrete realities. It is charged that most laboratory experiments have to do with the analysis and description of such abstractions as sensation, perception, memory, and imagination, with little or no attention paid to the way in which these processes are integrated into the behavior of the total organism. The objects of study are parts of individuals—not whole individuals.

While it is true that science must proceed by analysis, yet the transition from analysis back to synthesis is always a difficult one. Analytic chemistry has reached a higher state of development than synthetic chemistry. It is very easy for a chemist to analyze a loaf of bread into its constituent elements, but no chemist has as yet been able to manufacture a loaf of bread out of these elements. In the same way experimental psychology has analyzed parts of individuals in great detail, has noted many elements, and described them with considerable minuteness, yet no psychologist has yet been able to piece together anything that resembles a mind, much less a human being, from the analyzed parts.

Experimental psychology, having during the past forty years established itself with reasonable security among the natural and biological sciences, is less constrained than it once was to confine itself to the areas of strict laboratory techniques. The movement today is decidedly in the direction of studying the individual as a whole. For example, Watson uses this as the chief differential between psychology and physiology. Physiology, he claims, deals with the functioning of the organs as parts; psychology deals with the functioning of the same organs from the point of view of the individual as a whole. Whatever may be the merits or demerits of behavioristic psychology, it certainly has emphasized the importance of the total individual in contrast with the abstracted and unrelated parts.

The Gestalt psychology is also a movement in the direction of synthesis. Whatever may be the merits or demerits of this movement, it is also in the direction of the study of the individual as a whole. The same may be said of Freudianism and the whole psychoanalytic movement. Still further, the practical branches of psychology, such as industrial psychology, business psychology, and the like, all place the emphasis upon the individual as an individual.

What is Individual Psychology? This question has already been partly, if not wholly, answered. The major problem of individual psychology is the ways and amounts of individual differences. The data or subject-matter of individual psychology are the facts resulting from the studies of individual differences. Inasmuch, therefore, as the problems and the data of this branch of psychology are individual differences, it seems best to change the name from "individual" to "differential" psychology, for in order to study individual differences there must be at least two individuals. If the individual is studied as an individual and as though he had no relation to other individuals, we would then be back in the region of traditional experimental psychology. Differential psychology studies the individual at the points where he differs from other individuals. This at once suggests that differential psychology is not very far removed from social psychology, because social psychology also studies individuals in their relations to each other.

What is Social Psychology? What are its data and problems, and how does it differ from other kinds of psychology? These questions are by no means easy to answer. Social psychologists (or, more accurately, persons who write books on this subject) do not agree in their answers to these questions. Three major points of view and methods of approach may be discovered in their writings.

First, there are those who differentiate social psychology on the basis of stimulus types. For instance, Allport (1) defines social behavior as "behavior in which the responses either serve as social stimuli or are evoked by social stimuli." According to this view, social psychology deals with that area of individual behavior which is the result of social stimulation or which in itself may serve as a social stimulus.

Secondly, other writers find the differential of social psychology on the side of the response. Social behavior is characterized not by the nature of the stimulus but by the nature of the response. One advocate of this view is Kantor (24, p. 433), who states his position as follows: "What then constitutes the difference between a social and an individual psychological fact? Precisely this, that the data of social psychology constitute cultural reactions, that is to say, reactions which are built up by a conformity process, a process, namely, by which we become adapted to institutions or cultural stimuli which consist of things or actions found in the group or groups in which we live. In this sense social psychology is always a genetic study."

Other advocates of this view insist that social behavior is differentiated from individual behavior by virtue of the fact that in group actions, as for example, a football team, the effect of the whole is greater than, or at least different from, the sum of its individual parts. This phenomenon of "emergence" is illustrated by Woodworth (41, pp. 177-178) in the following way: "Suppose we have three dots on the blackboard, arranged in a particular form, triangular, let us say. Then it is perfectly true to say that the dots are all there is there, except indeed the homogeneous spatial medium. Where then is the triangular form, since it certainly does not reside in any one of the dots taken alone, nor in the three, if each is regarded as isolated from the others? The triangular form resides in the arrangement and mutual relations of the dots. These are purely static relations, but if we consider the actions of things, we find similar cases of dynamic relations and patterns. . . .

In the same way, when two or more human individuals are together, their mutual relationships and their arrangement into a group are facts which would not be disclosed if we confined our attention to each individual separately; and, when they act together, upon some common object —which may be one of themselves, or some other person, as well as a non-human object—the combination of their actions is a fact that could not be observed by considering the individuals one at a time."

A third point of view is found in the writings of such social psychologists as McDougall (27), who find the differential in neither the stimulus nor the response but in the driving force. The question posed by this point of view is not, "What constitutes social behavior?" but rather, "How does it happen that individuals get together in groups at all and how may such behavior be explained?" There is no general agreement between those who approach the problem from this point of view. For example, long ago the philosopher Hobbes took the position that individuals group themselves into communities and societies as mutual protective associations, their basic instinctive drive being fear. Tarde, Baldwin and others found the explanation in suggestion and imitation. McDougall finds it mainly in a set of seven instincts with special emphasis on self-assertion and submission.

Can These Points of View be Merged? Various attempts have been made to synthesize these three points of view. Such attempts usually end with efforts to distinguish between the individual and the group. The essential problem is to find out what is a group and what is an individual. This is the main concern of this chapter. The sociologists and social psychologists have endeavored to make this distinction clear. It is pointed out that when we contrast the individual on the one hand with the group on the other hand we create a false antithesis. The problem of the relation of the individual to the group is essentially the same as the age-old debate, "which existed first, the egg or the hen."

Cooley (11, pp. 1-2) has discussed this matter at great length. One paragraph is worth quoting: "A separate individual is an abstraction unknown to experience, and so likewise is society when regarded as something apart from individuals. The real thing is Human Life, which may be considered either in an individual aspect or in a social, that is to say general, aspect; but is always, as a matter of fact, both individual and general. In other words, 'society' and 'individuals' do not denote separable phenomena, but are simply collective and distributive aspects of the same thing, the relation between them being like that between other expressions one of which denotes a group as a whole and the other the members of the group, such as the army and the soldiers, the class and the students, and so on. This holds true of any social aggregate, great or small; of a family, a city, a nation, a race; of mankind as a whole: no matter how extensive, complex, or enduring a group may be, no good reason can be given for regarding it as essentially different in this respect from the smallest, simplest or most transient."

Ogden (30, pp. 89-90) has summed the matter up rather neatly: "There are two different conceptions of society. According to one of them a society consists of members which are 'integrals' in the group. According to the other conception society consists of separate 'integers,' each a self-sufficient individual. In the first conception the group-whole is more significant than the individuals that compose it; whereas in the second the group, being constituted of individuals, is less significant than its members. In choosing the first of these two conceptions we maintain that a social group determines its members, each of whom will possess whatever degree of individuality that the particular group makes possible . . . Social integration is not the summation of 'integers'; instead, the members of a social group are the 'integrals' of a whole. In other words, priority attaches to the group and not to its members."

Following the lead suggested in this quotation it would appear that social psychology may be subdivided into differential and integral.

It is proposed then that we distinguish between "differential" social psychology and "integral" social psychology. Now the data of differential social psychology are the facts of individual differences and the problem is to determine the ways in which individuals differ, the amounts of these differences, and the social significance of the amounts. The data of integral social psychology, on the other hand, are the facts of group conduct (meaning the behavior of individuals in groups), social institutions, customs, traditions, taboos, and the like. Its major problem is to explain the phenomena of social integration, that is, how it happens that individuals work together as a whole in teams, communities, societies, states, and nations.

The place of the adult in the community cannot be determined until we find out the ways in which "our hero" differs from other adults and the amounts of these differences. When these determinations have been made, the next problem is to find out how he functions in his community, not merely how he is adjusted to it.

Differential Social Psychology

The procedure in finding out how individuals differ is exactly the same in psychology as in chemistry. Differences between substances or elements are obtained by subjecting them to the same conditions and noting the results. When two substances react differently to the same conditions they are different; when two individuals respond in different ways to the same stimuli they are different individuals. Thus, when an individual behaves in a manner that deviates from or is different from his accustomed modes of response, we say that he is not himself.

If all individuals were the same in all respects, the world would not only be a dull place, but there would be no social integration—only summation. A shoal of fishes or a migratory flock of birds represents a loose formation of social organization because there is little or no differentiation. For this reason it is important that social psychology look for its factual data in the behavior of individuals, and especially in their differential behavior. Since differential psychology has not been treated as a part of social psychology it will be emphasized in this chapter more than the integral aspect. We shall begin by considering the biological background of differences, follow this through to differences in animal societies, and finally consider some of the important individual differences among humans.

The Biological Background of Differentiation

Differentiation in lower animals. The biological problem is: How does it happen that all individuals, even in a given species are different? Why is it that no two animals are exactly alike, no two leaves on the same tree exactly alike, no two anything, in fact, in nature exactly alike? Two theories have been advanced to explain this. One emphasizes the hereditary aspects of the situation, the other the environmental aspects. It is more accurate to say, however, that there are two kinds of theories, because among the hereditarians we find a great many sub-theories, and among the environmentalists many sub-theories. This is not the place to discuss the facts and the merits of the two sides of the argument. It is enough to say, however, that the modern tendency is toward conciliation. Instead of saying that it is all a matter either of heredity or of environment, the question today is how heredity and environment pull together in producing variation.

A few years ago when biologists and geneticists were concerned with the morphological aspects of animals, the heredity hypothesis seemed to fit the facts best, but now the biochemists and physiologists and biologists have turned their attention to the functional and behavior aspects of the organism, and the environmental hypothesis seems more appropriate. As the case stands today, it appears that there can be no doubt concerning the inheritance of protoplasmic structures. On the other hand, there is not much evidence in support of the theory that specific behavior patterns are inherited except in so far as they depend upon structures. Each organism probably inherits a general behavior pattern that fits its structure. The formula seems to be somewhat as follows: heredity sets the limits or determines the general nature and range of the behavior patterns, but the patterns that are actually developed in any individual are the results of the interplay between the organism and its environment. Thus, individuals of the same species may have the same heredity potentialities, but the range or scope of these potentialities in any individual is usually so much larger than what is realized in the development of the individual that it is quite possible for all individuals of the same species to be different. On the other hand, it is quite possible that the limits and scope of potentiality set by heredity are so different that two individuals could not possibly become the same, regardless of the most constant environmental influences.

If it is true that no individual ever realizes all of his hereditary potentialities, and if it is true that those potentialities realized by any one individual are determined by the environmental interplay within the limits set by heredity, how does it happen that environment produces individual differences in behavior? In general the impact of environmental energy upon the organism influences it in two ways: first, the organism adjusts itself to these impacts; and, secondly, it sets up a machinery for effectively controlling its environment. A double adjustment takes place. In order for the organism to adjust itself to its environment or to adjust its environment to itself, differentiation of parts is necessary. In order to understand how individuals differ from each other it is necessary for us to understand how one part of an organism comes to differ from the other parts.

The Differential Effects of Environment. One hypothesis as to how differentiation takes place within an organism is the "physiological gradient" theory of Child (7). The physiological gradient is essentially a differential rate of living. Physiological activity, including chemical and physical velocities, colloid dispersion, and ionization, takes place with different rates and intensities in different parts of the organism. Child finds abundant evidence for supposing that these activities are graded quantitatively from an area of maximum to areas of minimum effects. When environmental energies impinge upon the organism, they set up certain activities within the organism and the point of impingement is for the time being the most effective region from which the activity spreads to other regions. As it spreads it fades and becomes less and less until it finally ceases. This inequality in effectivity of rate of living is called a "physiological gradient."

The hypothesis advanced by Child is that the structure and functions of the different parts of an organism are determined by such gradients. He cites experimental work showing that by controlling the gradient it is possible to develop organisms with small heads, dwarfed brains, or with the eye of a Cyclops located in the median plane. By modifying or completely obliterating, or by introducing new gradients, he is able to change the structure and functions of an organism in a great many ways. For example, the characteristic structural form of polarity and symmetry of many organisms may be completely obliterated, leaving the organism spherical, by the proper control of the determining gradients. "In short," says Child (8, p. 27), "by controlling the physical or chemical environment of the average individual it is possible to alter the whole pattern to such an extent that the resulting individuals differ far more from the normal than even widely separated species do from each other; moreover, these alterations are not random occurrences but can be predicted to a high degree."

Child vigorously defends the thesis that the individual takes on individuality by virtue of his contacts with environment. Differences in the behavior of different members of the same species are due mainly to the different effects of the environment on the protoplasmic organism. In lower organisms wide varieties of individual differences may be obtained under experimental conditions, which illustrates that no individual realizes more than a fraction of its heredity possibilities. Furthermore, in no two individuals is the realization the same, since they do not except by chance come under identical environmental stimulation; "the hereditary constitution of each protoplasm reacting to the external differential gives first the physiological gradient and then, on the basis of this, development and differentiation" (7, p. 134).

It should be noted that this view combines in an effective manner the hereditary and environmental hypothesis. Heredity supplies the potentialities of excitation and reaction; it also determines the degree of persistence or effectiveness of the gradient. Environment, on the other hand, determines which gradients shall be operative and how they shall operate. Thus it may be seen that the view of Weismann and others, who claim that individual differences are due to chance variations in the determinants contained in the germ-plasm, is gradually losing ground.

Differential Animal Psychology

Differentiation in higher animals. Animal psychology has paid more attention to the ways in which animals are alike than to the ways in which they are different. The descriptions of animal behavior refer in general terms to a whole species, with very little reference to how the different individuals of the same species behave in the same environment. One reason why individual differences in lower animals has been a neglected topic is that each animal has been supposed to behave in a completely stereotyped manner. For example, certain species of bees, wasps, and ants have been supposed to possess fixed methods of nest-building, brood-production and food-gathering. Deviations from these methods were regarded as the exception rather than the rule. Recent studies of insects have revealed that their modes of behavior are by no means as fixed as was supposed. Their environments do, as a matter of fact, have differential effects on them.

Environmental adaptations. The fact that different members of the same species adapt their activities to their particular environment has often been noted but has not been treated from the point of view of individual differences. It is well known that spiders of the same family do not always spin webs that are geometrically identical, but each web is modified by the pattern set by the objects to which it is attached. Birds of the same species do not always build identical nests but use such materials as they can find, which they adapt to the shape of the crotch of the selected tree. The female *Ammophila* wasp, made famous by Fabre for her cunning trick of paralyzing a caterpillar by stinging it in a certain nerve ganglion and carrying it to her nest paralyzed, but not dead, and storing it there for her brood so that the baby wasps will have fresh meat upon which to feed when they are born, is supposed to be an example of unerring accuracy of instinctive patterns. But when Mr. and Mrs. Peckham studied this behavior in detail they found that the activity is not as accurate as advertised. Apparently the wasp stings the caterpillar by a series of upward thrusts which often go wide of the mark. An examination of the caterpillars in wasps' nests shows that many of them are really dead. No doubt if each female wasp could be subjected to some kind of an efficiency test there would be wide individual differences in the ability to hit the bull's-eye.

Individual differences in the quality of song birds have long been observed by naturalists. Chaffinches sing more beautifully in some sections than in others; nightingales sing better in communities where they are protected from their natural enemies than in places where they must be on their guard. Naturalists have observed that older birds teach their young ones to sing, or at least the song qualities of young birds are greatly modified by hearing the older and better singers. This observation has been experimentally verified in experiments with Baltimore orioles. Some orioles were isolated when they were very young without an opportunity to hear the song of a bird of their kind. They became good singers, but their songs did not resemble that of an oriole. Scott found that he could modify the songs of different kinds of birds by the kind of singing they were permitted to hear. He found, for example, that catbirds can imitate the postman's whistle, and that one red-winged blackbird actually learned to crow like a bantam rooster. Here again, if some sort of a singing test could be devised that would measure the quality of music produced, there is no doubt that birds of the same species would exhibit wide individual differences.

From these facts it is evident that animals do not behave according to a rigidly cut-and-dried fashion. Heredity provides the broad patterns, limits the possibilities, and determines the potentialities which are realized in different ways by different individuals.

Individual differences in animal societies. Further light on the process of differentiation is obtained by the study of the more integrated animal societies. Certain species of bees, wasps, ants, and termites have been studied carefully and much is known about their habits and forms of organized life. One of the outstanding characteristics of the social insects, and one that has amazed naturalists, is their plasticity, i.e., their capacity for differentiation. This is especially true of ants. They are plastic and variable in their nesting or home-building habits. Whereas the wasp and the spider seem forever doomed to a certain nest structure, the ants vary their homes, as do men, to suit the conditions of temperature, moisture, or the habits of their enemies. This plasticity is not only confined to nest building but extends to care of the young, feeding habits, fighting, and many other enterprises. "All this adaptability, or plasticity in nesting and feeding habits," says Wheeler (39, p. 154), "is, of course, an expression of a very active enterprising disposition and has resulted in the formation of a vast and intricate series of relationships between ants and other organisms, including man. These restless, indefatigable, inquisitive busybodies, forever patrolling the soil and vegetation in search of food, poke their noses, so to speak, into the private affairs of every living thing in their environment."

In all ant societies there is always a queen or an egg-layer, who does nothing but just that. Next in importance are the workers, who may or may not be further subdivided into castes. In some species of ants this subdividing of the workers is very detailed and minute. Some of them have as their duty that of cleaning and attending the queen; others feed and clean the larvae; others tend the fungus gardens; others go foraging for food; others repair or extend the nest; others tend the aphid cows; and so on for a variety of duties, depending on the species and on the size of the colony. Another major division is that of the army, and here there are soldiers and sentinels, and, if a recent newspaper report is true, some colonies have a fire department. In short, the divisions of labor follow roughly the lines of the four main necessities of life, which are reproduction, food, shelter, and protection from enemies.

Little seems to be known concerning the tenure of office in these various castes in ant societies. It is commonly assumed, that if an individual ant is once a soldier he or she is always a soldier; once a gardener, always a gardener. It would be indeed interesting to find if there is any such thing as promotions among them so that an ambitious chambermaid, for example, could work up to the post of corporal of the guard. It is known, however, that in some species jobs are delegated according to the size and structure of the individual, indicating that here at least some sort of personnel work is carried on. Thus Alverdes (2, p. 91) observes, "The division of labor is, in some species, connected with a morphological polymerism in the workers; two, three, or more types are found. Thus in the case of *Atta,* the middle-sized workers are leaf-cutters, the large ones leaf-grinders, and the smallest tend to fungus-gardens. Soldiers, again, are of a special type." This would indicate that differentiation of functions is in part a matter of structure. But it is not wholly so. Differentiation may be on other grounds. For example, some may secrete chemical substances that others do not; some may be more efficient than others. What are the facts on how differentiation or division of labor takes place? What determines the place of the adult ant in his community? Wheeler (39, p. 161) thinks that "in the present state of our knowledge we can only surmise that the differences between the queen and worker castes were originally ontogenetic and determined by feeding, as they still are in the social wasps and bees, but that in the ants the germ-plasm has been somehow reached and modified, so that an hereditary basis for caste differentiation has been established."

In certain animals, however, more evidence is at hand on how the position of the adult in the group is determined. One of the most interesting is a study made by Schjelderup-Ebbe on barnyard hens. When a flock of hens are placed together in a run, a "pecking order" is soon established by fighting. When a newcomer is introduced, her position on the "pecking list" is soon established. The "pecking order" may take various patterns. Sometimes it is hierarchic and graded so that the superior hen pecks on all the others; the next in line pecks on all but the one above her, and so on down to the last unfortunate hen who is pecked on from all above and has no one below her on whom she may peck. Other patterns may be triangular, in which A, B, and C peck in the order: A on B, B on C, and then, surprising as it may seem, C on A. Or again, a quadrilateral order may be observed in which A pecks B, B pecks C, and C pecks D, and D pecks A. Such orders are partially explained by the fact that some hens will submit without fighting. For example, C may submit to D without fights, but D wins over A. Similar pecking lists are established among roosters, sparrows, and wild ducks. The position of any fowl on the list is not determined by strength or fighting efficiency alone. Other factors enter in, such as time of arrival with the group, age, and even intelligence.

Facts such as the above will serve to indicate that in animal societies certain degrees of differential functions and social positions are present. The more complex the society and the more integrated it appears, the more diverse the individual differentiation between members. This fact is indeed almost axiomatic, because without diversified individuation there can be no integration.

Differential Adult Psychology. Human adults differ from one another in countless ways. In fact, it is doubtful if there is a single anatomical feature, or physiological function, or psychological process in which all individuals of the homo sapiens are identical. Two kinds of differences are recognizable, qualitative and quantitative. In gross physical anatomy and physiology and in certain gross psychological processes all humans are qualitatively alike. This simply means that they all have heads, hearts, lungs, stomachs, and all hearts pump blood, all lungs breathe air, etc. Moreover, their psychological likenesses are of the nature of such broad features as memory, imagery, sensitivity to stimuli, ability to make muscular responses, reflexes, and the like. But within each and every one of these qualitative likenesses there are vast quantitative differences. There is no feature of the human anatomy or physiology nor of mental processes in which all individuals are identical. The dictum that everything that exists, exists in amounts, applies to human abilities, capacities, and anatomical structures as well as to facts in the inorganic world.

The importance of this fact for social psychology has long been recognized, but like many other obvious and commonplace phenomena, it was not made the object of scientific study until the last quarter of the 19th century. The first scientific studies of individual differences were made by Sir Francis Galton, an eminent English scientist. Galton was not a psychologist, at least he was not a member of the profession as it existed in his time, but a traveler, explorer, and a scientific adventurer. Influenced, no doubt, by his illustrious cousin, Charles Darwin, he set out early in his career to study the human species in much the same fashion as the biologists of his time were studying animals.

Galton is distinguished as being one of the first to sense the importance of individual differences for social organization and integration. During his time England was struggling with one of the most stupendous problems of social integration that has ever faced human society—that is, colonization. In the introduction to the second edition of his *Hereditary Genius* he says: "My object now is merely to urge inquiries into the historical fact whether legislation, which has led to the substitution on a large scale of one race for another, has not often been the outcome of conflicting views into which the question of race hardly entered at all, and which were so nearly balanced that if the question of race had been properly introduced into the discussion the results might have been different. The possibility of such being the case cannot be doubted, and affords strong reason for justly appraising the influence of race, and hereafter including it at neither more nor less than its real value, among the considerations by which political actions will be determined" (17).

In his volume on *Inquiries into Human Faculty* we find the same interest in fundamental human problems. He opens with statements such as these: "My general object has been to take note of the varied hereditary faculties of different men, and of the great differences in the different families and races, to learn how far history may have shown the practicability of supplanting inefficient human stock by better strains, and consider whether it might be our duty to do so, by such efforts as may be reasonable, thus exerting ourselves to further the ends of evolution more rapidly and with less distress than if events were left to their own course" (18, p. 1).

"The moral and intellectual wealth of a nation largely consists in the multifarious variety of the gifts of the men who compose it, and it would be the very reverse of improvement to make all of its members assimilate to a common type" (18, p. 2).

During the first quarter of the present century an army of workers have been busy studying individual differences. Human adults and children have been studied in almost every conceivable way. These investigations may be roughly divided into qualitative and quantitative. The qualitative studies have been concerned with identification and isolation of traits or variables in which individuals differ. The quantitative studies have been concerned with the measurement of these differences. The variables which have been isolated and measured are too numerous to be listed here. The best that can be done is to classify them and give a few examples. For convenience they are divided into four main heads.

Table of Variables in which Individuals Differ

1) *Anatomical structure*

a) *Gross physical features.* Here are included individual differences in such things as height, weight, length of different bones of the body, size and shape of the head, and other gross anatomical features. All sorts of body measurements have been taken and various ratios have been worked out. Two of the best-known ratios are: first, the cephalic index (ratio of length to width of head); and second, the morphological index, which is the ratio of height to general width or size of the body. This index has been variously expressed.

b) *Finer anatomical measurements.* Here we have such things as individual differences in color of eyes, texture of skin, peculiarities of skull, number and arrangement of teeth. Many years ago when phrenology and physiognomy were rated as of great social importance, much time was devoted to the study of what one might call cranial topography and facial geography.

2) *Individual differences in physiological variables*

a) *Gross physiological functions.* The variables here that have received most attention have been such things as physical strength, including strength of grip, strength of muscles of the legs and the back, and other parts of the body, variations in physical endurance, susceptibility to fatigue, lung capacity, strength and efficiency of the muscles of the circulatory system, especially of the heart, gross motor control, motor coordination, and speed.

b) *Finer physiological functions.* Physiologists have succeeded in isolating a great many variables in which individuals are known to differ. Some of these are: rate of basal metabolism, speed of nerve conduction, the chemical constituents of gland secretions, hydrogen-ion concentration in certain fluids of the body, degree of acidity in certain fluids of the body, and others.

3) *Psychological variables*

a) *Sensory.* Individual differences in sensory acuity were among the first types to be studied. These differences occur mainly in the range of effective stimulation (sensory limen) and in ability to distinguish between the quantitative differences in the same type of sensory stimulus, as, for example, individual pitch discrimination, or individual differences in the discrimination of lifted weights, and the like. This particular field has been developed to a high degree of precision by a sub-branch of psychology known as psychophysics.

b) *Individual differences in motor variables.* Some of these have already been mentioned under physical variables. One of the first of these to be measured in psychology was reaction-time. Following that, other sorts of time variables were measured, such as association time, discrimination time, and the like.

Individual differences in the finer motor capacities, such as speed of tapping, motor steadiness and control, have been determined.

c) *Instincts and emotions.* Individual differences in strength of instinctive drives, strength and variety of emotional responses, have been studied, though not very accurately measured. All endeavors to measure the emotional variables have usually resulted in measuring certain of the finer physical functions.

d) *The so-called higher mental processes.* Here we have a vast array of variables, such as differences in all kinds of mental abilities, differences in a great variety of imageries, differences in association, and the like. In fact, each of the traditional psychological categories has been analyzed in search of variables that could be measured quantitatively.

e) *More generalized capacities.* The rise of intelligence tests and the success that has attended the efforts at measuring general intelligence have encouraged psychologists to seek for general traits or general abilities that can be measured on a single quantitative scale. The most important of these more generalized functions that have been measured with a reasonable degree of success are intelligence, musical ability, and mechanical ability.

4) *Integral variables.* Individuals also differ in the degrees of organization and coordination in the foregoing list of variables. They also differ in social adjustment and in their social integrations. Such differences are usually called character and personality. Quite recently efforts have been made to isolate certain character or personality qualities that can be measured quantitatively. Measuring devices have been invented and certain data are available on such traits as ascendance-submission, confidence, honesty, persistence, suggestibility, and unselfishness. Further efforts have been made to measure such generalized abilities as social attitudes, range of interests, strength of prejudices, and moral judgments.

This survey of some of the variables which have been measured or isolated for measurement is more schematic than complete. It is not intended to serve as an entire catalogue but merely as an illustration of the kinds of variables in which individuals have been found to differ. To report the results of all of the studies of these individual differences would be indeed a foolhardy undertaking. The best that can be done is to select as illustrations certain variables in which individual differences appear to have the greater social significance.

Social Significance of Individual Differences in Anatomical Variables

Individual differences in physique. As illustrating the part played by certain anatomical features in determining the position of the adult in the community, we shall consider *physique* and *physiognomy.* On the side of general physique we have already noted that in certain animal societies gross bodily size seemed to be a determining factor in social differentiation. Division of labor seemed to be made mainly according to the bodily structures of the individuals. The larger individuals were given jobs as soldiers and sentinels, whereas the smaller ones were assigned duties such as house-work and work in the nursery. In human societies we still see that bodily size plays some part in social differentiation. Policemen, athletes, and soldiers are selected in part at least on the basis of their physical prowess. The extent to which mere physical size or strength is a social asset depends on the community, the occupation, and, to some degree, on the state of civilization.

Gowin (19) tabulated the heights and weights of eminent persons, such as United States senators, governors, bishops, presidents of industries, and others. The average height of 1,037 such individuals was found to be 71.4 inches and the average weight, 181.1 pounds. Gowin compares these averages with those of 221,819 applicants for life insurance as tabulated by insurance companies. These applicants have an average height of 68.5 inches and an average weight of 166 pounds. Thus, the average height of the 1,037 eminent men is nearly 3 inches greater than that of the life-insurance applicants, and the average weight 15 pounds greater than the average weight of the life-insurance applicants. From this study Gowin concludes that, on the average, men who are big in the affairs of the country are bigger physically than the average.

While the scientific studies of the relation of physique to social adjustment are by no means satisfactory, yet common observation leads to the suggestion that individuals who are above the average in height and weight do, as a matter of fact, make different adjustments from those who are below the average. It has been pointed out many times that individuals of small stature tend to develop compensatory mechanisms. Undersized individuals and those of short stature compensate by making plenty of noise and by standing heavily on the ground. The oversized person, on the other hand, is likely to compensate by his apparent docility and good humor. These, however, are only commonplace observations and need to be checked by accurate counts and measurements.

Kretschmer (26) has recently made a heroic attempt to hitch up the two main psychiatric types with certain types of physique. He studied the physiques of 85 individuals of the manic-depressive type (circulars or cycloids, as he calls them) and 175 cases of the dementia precox type (schizophrenes or schizoids, as he calls them). He found three rather distinct bodily types; first, the *asthenic,* characterized by lean, narrow, and thin body, pale and angular face, sallow complexion, and flat chest; second, the *athletic* type, characterized by broad shoulders, muscular development, thick skin, rugged face, and of medium height; third, the *pyknic* type, characterized by round flat figure, soft, broad face and massive neck, well-developed abdomen, and soft delicate hands. His deduction is that the manic-depressives belong for the most part to the pyknic type, and the dementia precox to the asthenic and athletic types. From this he proceeds to hitch up various temperaments and occupational inclinations with bodily types. For example, among leaders he finds that the pyknics are tough, hard-boiled whole-hoggers, while the asthenic and athletic types are the idealists.

Kretschmer's studies are attracting a great deal of attention and have enticed many other investigators into this field. All sorts of studies have been tried in an effort to find relationships between physique and character. Naccarati, following Viola and the Italian School, determined the morphological index (roughly, the ratio of height to weight) of adult students and correlated the results with their intelligence. He found low but positive correlations. This work has been repeated and extended by others, and for the most part the results are verified.

The case then concerning the relationship of individual differences in physique to social adjustment is about as follows. When normal individuals are taken as subjects, it is very difficult to establish correlations between any measurable anatomical variables and any measurable social variables. When extreme cases of social maladjustment are taken, as, for instance, the insane, then certain types of bodily structure do seem to fit with certain types of social maladjustment. These facts indicate, then, that physique does play some part in determining the position that the individual shall occupy in his community, but how great this part is remains for science to determine.

Individual differences in physiognomy. Approximately one hundred years ago phrenology was at its height. It was believed that there were close correlations between certain character and personality traits and certain gross geographical features of the skull. This, however, has been rather generally discarded, and there has grown up in its place another pseudo-science known as physiognomy. Physiognomy does not place major stress on "bumps," but rather on the general facial appearance of the individual. The important variables are such things as width between the eyes, height of forehead, color of eyes, color of skin, length of nose, and the whole facial contour. The chief exponent of physiognomy is Dr. Katherine Blackford, who has worked out a whole system for determining social traits by means of these anatomical structures. Two or three scientific investigations have been made which show that the correlations between cranial topography and facial geography, on the one hand, and social traits, on the other, are by no means as close as advertised. For example, Paterson and Ludgate (31) selected 374 cases of blondes and brunettes who were rated by 94 competent acquaintances on each of the Blackford traits. They found that the brunettes possessed blonde traits in the same proportion as the blondes, and vice versa. Cleeton and Knight (10) worked on the same problem but approached it from the other side. They selected individuals who seemed to possess marked degrees of traits that are supposed to have anatomical correlates, but when they made precise physical measurements they found that these individuals did not have the anatomical structures that were supposed to go with their traits. For example, sound judgment is supposed to be indicated by wide distance between the eyes. They took those individuals who were rated by their acquaintances as having the soundest judgment, and measured the distance between their eyes. They took the individuals who were rated by their acquaintances as having the weakest judgment, and measured the distance between their eyes. The results came out the same for both groups. This process was repeated for other traits. The results were rather uniform in showing that there are very few, if any, genuine correspondences between social traits and anatomical variables. Such studies, however, are open to criticism on the ground that the anatomical features were measured rather precisely and objectively, while the social traits were rated subjectively and were open to certain errors. When psychology has reached the point where social traits can be measured as objectively and accurately as anatomical features, then and only then can this question be decided.

Race differences mainly anatomical. The greatest social significance of individual differences in anatomical structure, such as color of skin and hair and shape of head, lies at present in the part they play in race differentiation. While it is true that from the biological point of view clear, clean-cut race differences are hard to find, and while it is true that most race differences are fundamentally differences in culture, yet the anatomical features of races, such as the Japanese and the negroes, are sufficient to mark them off and identify them. Indeed, it is true that the only sure way of determining the racial affiliation of any individual is by some anatomical or physical trait, and even here such traits are not always accurate. It is exceedingly difficult to identify the race of an individual by conduct, ideals, attitudes, or any psychological peculiarities. Suppose, for example, that an individual is described as being stingy, clannish, able to drive a hard bargain, average in intelligence, rather devoted to the church and religion; would it be possible to name the race of this individual from such a description? This description would fit very well either the Scotch, the Jewish, or the Armenian, and perhaps even the Greek. Suppose now, on the other hand, that we describe an individual as having a yellowish brown skin, black hair, black eyes, five feet two inches tall, weighing 135 pounds, male, age 35, thin lips, with eyes slanting in and downward, round face, with forehead of medium height. What is the race of the individual? No one would doubt that this individual is either Chinese or Japanese, probably Japanese.

Thus we see that it is much easier to identify race's from an anatomical description than it is from a psychological or sociological description. We venture the guess that race is mainly a matter of anatomy. If by some miracle all the negroes in the United States should suddenly turn white, their hair become straight, and their lips thin, the negro problem as a social problem would at once disappear. On the other hand, if all white people should suddenly turn black, their hair take on a permanent wave, their lips become thick and their noses flat, the race problem would also disappear. It is clear that individual differences in color of skin has very far-reaching social significance. That there are race differences in customs, dress, personal habits, sanitation, and the like, there can be no doubt, but when it comes to intelligence, temperament, and skills, genuine race differences are hard to find. The one trait on which there is the most data is intelligence. Race differences in intelligence have been discovered, but there is considerable debate concerning their genuineness and their social significance. There is enormous overlapping among the races in this, as in every other psychological trait, in which individual differences have been measured. It is true that there is more variability within any one race in respect to psychological capabilities than there is among the average scores of the different races. The point is that there is no single psychological trait or set of psychological traits which, when measured by the most accurate instruments known to science, will show scores by which a member of a given race can be spotted as surely as by gross anatomical features.

Social Significance of Individual Differences in Physiological Variables. Here we shall take sex differences as an illustration.

The social significance of sex. As a physiological function sex cannot be called a variable. It is one of the few functions that permits of a definite two-fold classification. The fact that the two sexes have occupied separate rôles in the process of reproduction has made it one of the most important factors in determining the position of an adult in the community. The position of the father and the mother in the family—the most elemental human community—is determined pretty much by the fact that the mother bears the children and feeds them while they are very young. There is nothing, however, in the physiology of the case to make it necessary for the mother to care for the child beyond the period of lactation. It is important to recall that the function of child-bearing and child-tending has become such an intimate part of the life of mothers that there have grown up a theory and conviction that this is their major duty and their major part in community life. This conviction has become so strongly intrenched that when women have come forward and asserted their "rights" there has come (mainly from men and some women) a howl of protest. When mothers "step out" of the home and take on other duties in the community, such as business, office-holding, and the like, it is viewed as the beginning of the disintegration of the family and the ultimate deterioration of the community. That the woman's place is in the home is a common and trite, yet firm belief of many.

Modern psychology is gradually accumulating a body of evidence on sex differences which shows rather conclusively two things: (1) that sex as a determinant of the position of an adult in the community is based largely on tradition, and (2) that apart from the reproductive functions there are no differences between the sexes that can be traced to fundamental hereditary behavior patterns. The facts concerning sex differences may be summarized very briefly.

There is no aspect of individual psychology that has received more attention than that of sex differences. From 1900 to 1928 more than 200 articles have apeared in psychological journals dealing with sex differences in mental functions alone. Add to this the articles and books on sex differences in other respects, and the total literature on the subject amounts to many hundred papers comprising thousand of pages. Out of all this have come the following general conclusions:

a) In respect to intelligence, or intellectual capacity as measured by a variety of intelligence tests, there are no consistent differences between the sexes when fair comparisons are made. By fair comparisons we mean comparison based on data in which the two sexes are equal in respect to age, nationality, and other variables that are correlated with intelligence. Evidence that no sex differences in intelligence exist is the fact that mental test builders no longer publish norms for the sexes.

b) In respect to variability. Finding no differences between the average scores of two comparable groups, it was suggested that possibly there may be a difference in the spread or variability of the groups. For example, it has been observed that there are more males than females in institutions for mental defectives, and also, at the other end of the scale, there are more males than females in the lists of eminent and accomplished persons. But careful investigations of this problem have resulted in the conclusion that this observation is due to selection. For instance, the fact that there are more males than females in institutions for mental defectives is explained by the fact that a defective female is more likely to be cared for in the home than a defective male. Mrs. Hollingworth has compiled data on some 2,000 mental defectives showing that "a female with a mental age of six years survives the social milieu about as well as does a male with a mental age of 10 or 11 years." The reason that there are more names of males on the lists of the eminent is that women have by social tradition been so cramped and restricted and confined to home duties that they have not had equal opportunities to rise to eminence. In support of this contention is the fact that during the past 25 years women have begun to assert their rights and step out from home duties to wider social activities, with the result that their roll call in *Who's Who* and other such dictionaries is increasing.

c) Females are supposed to excel the males in linguistic ability, the manipulation of language; while boys are supposed to excel the girls in mechanical ability, the ability to manipulate tools. Passing over the tradition that women are more "talkative" than men, and confining ourselves to more scientific studies, we find that school marks do show that girls receive on the average higher grades in "language subjects," and boys slightly higher grades in the "science subjects." But no one has been able to show that this difference is one of ability rather than of interest. Actual tests of certain linguistic and mechanical abilities show no differences of great significance. Intelligence tests that depend on language manipulation (language tests) show no differences in favor of girls; nor do the performance intelligence tests show differences in favor of the boys.

d) In gross anatomical structure and in physical strength males are somewhat superior to females. Adult males are, on the average, somewhat taller and heavier than females; they also possess greater muscular strength, but, curiously enough, in tests of endurance the women tend to excel. The greatest differences occur in rate of growth and age of maturity. It is definitely established as a fact that girls grow faster and mature earlier than boys (that is, they reach puberty earlier). Other differences are of such order as differences in weight of brain, in which it has been found that the average weight of the male brain is about 1,400 grams, and of the female, 1,250 grams—facts subject to further revision. Yet no one has shown that the male brain possesses a greater number of neurons than the female, which, after all, is the important consideration.

e) In the finer physiological functions, such as glandular secretions, we find certain differences between sexes, especially in the action of the sex glands. It has been shown, with a reasonable degree of certainty, that the activities of the sex glands are associated with certain types of behavior characterized as masculinity and femininity. For example, the destruction of the ovaries in the female often results in the loss of certain feminine characteristics. In respect to other glands less is known. Thyroid diseases are more common in women than men, suggesting certain sex differences in the importance of this gland. Beyond this, little has been established concerning the differential effects of glandular secretions.

f) In certain personality and character traits all sorts of sex differences are supposed to exist. To mention a few—men are supposed to be more pugnacious, more self-assertive, more objectively minded, less introspective, more interested in things (women being more interested in persons), less excitable, less sensitive to social taboos, less inhibitive, and more emotionally stable. That general differences of this order do exist there can be no doubt. The questions are: how great are they, and what is their social significance? Due to the lack of accurate measuring devices, the amount of such differences has not been determined, but, since they are large enough to be commonly observed, we should assume that they would show up if measured accurately. Thus, we shall assume their existence and discuss their significance.

Their significance can be determined only in the light of their origin. Psychologists are inclined to believe that such differences are not due to differences in hereditary behavior patterns, but rather to the differential effects of environment. Men and women differ in their social attitudes and adjustments because they have had fundamentally different types of education. Parents and other guardians of social tradition have certain preconceived ideas of the sort of thing that is becoming or unbecoming to a girl or boy. For example, the little girl is supposed to be a thing to be petted, loved, dressed, and otherwise ornamented. Her life is cramped and limited in countless ways, her activities are laid out for her, her thinking is done for her, and her problems are solved in advance. Of course she is more interested in persons than things, because it has been her lot to deal with persons; of course she is less pugnacious than boys, because she has never been allowed to fight. She is less self-assertive because self-assertion is supposed to be unbecoming to a girl; in short, she is educated to occupy the traditional place of women in the community, namely, that of child-bearing and home-making.

In showing that there are no fundamentally innate sex differences in abilities and capacities psychology has done much for women. While the modern feminist movement originated elsewhere, it has been amply supported by the facts of individual psychology. Now that women know their capacities and abilities, they are asserting themselves quite vigorously, and in this their success is marked. They have gained the right to vote and to hold office and assume positions in the community on the same footing as men. The war helped them, too, for it gave them the opportunity to show what they can do. So today the position of the woman in the community, especially in America and England, is quite different from what it was one hundred years ago. This position, we repeat, has been determined by the recognition of the fact that the major difference between the sexes is physiological and not psychological.

Social Significance of Individual Differences in Psychological Variables. Here we shall take three illustrations, one from the field of sensation, one from the field of motor response, and one from the field of intellectual ability.

Individual differences in audition. Seashore (54) has made an elaborate study of individual differences in the three major auditory variables, which are pitch, intensity, and sense of rhythm or time. Seashore reports that some persons have 200 times as high ability in distinguishing pitch as others. In this variable alone, individual differences are probably as great as they are in any psychological variable. Wide individual differences are also found in the ability to distinguish sounds of different intensities, as well as in a sense of time and rhythm. These differences, along with others, are of great importance in determining an individual's musical ability.

The economic importance of such determination may be seen by considering the fact that the estimated cost of musical education in America is $220,000,000 annually. It has been estimated that the annual cost of musical education is four times as much as the cost of American high schools, three times as much as colleges and universities, and 24 times as much as normal schools; in short, we are spending more money on musical education than on any other kind of education in America. How much of this is wasted on people who have little or no talent is not known, but is undoubtedly very great.

The scientific attempt to measure the psychological variables that underlie musical ability has been crowned with marked success. This work has been carried on in the University of Iowa by Professor Seashore and his students for many years. The measurement of these variables has enabled those interested to spend their time, money, and effort in directions that will count for the greatest happiness and success of the individual and for the promotion of musical activities in the country. This study is an excellent illustration of the social significance of individual differences in sensory abilities.

Individual differences in reaction-time. The second illustration is taken from the field of motor response. One of the first and most clean-cut records of the significance of individual differences is the case of the observer at the Greenwich astronomical observatory in England who, in 1795, lost his position because he had a slow reaction-time. It came about in this way. The observer watched the star through the telescope and noted the time at which it passed a certain line. The procedure was that the observer should keep his eye on the star until it almost reached the line, then look at the clock and estimate the exact time when the star crossed the line. This particular fellow differed very widely in his estimates from those of his colleagues, and for this reason was discharged. It was discovered, however, some time later that the difference was due to his slow reaction-time. It was not, however, until 1882 that astronomers learned to allow for differences in the reaction-time of observers.

The social significance of individual differences in reaction-time has been studied in connection with other occupations besides that of astronomical observers. One of the more recent and one having great social significance is that of reaction-time of taxicab drivers. Among the tests that have been recently devised for the selection of successful cab drivers is that of Wechsler. Wechsler's chauffeur test consists mainly in determining the speed with which the candidate reacts to the flashing of lights of different colors at irregular intervals. He is told that certain lights will flash at certain times, serving as signals to which he must react in in certain ways by performance of the hands and feet. His responses are automatically recorded in such a way that individual differences can be tabulated. The results of such experiments show that, on the whole, quick reaction-time is one (but only one) important factor in the ability to drive a cab or a car with a low percentage of accidents. Other studies of reaction-time have shown that it has certain significance in other directions, but this one case will serve to illustrate the point.

Social Significance of Individual Differences in Intelligence. Intelligence is a complex affair. Psychologists have not come to any very definite agreement concerning its precise nature. To one thing, however, they all agree, and that is to its quantitative nature. It was Sir Francis Galton, who first suggested that intelligence could be measured by finding the place of the individual on a scale running from the intelligence of the lowest idiot to that of the greatest Athenian scholar. It remained for Binet to develop the first scale. Binet's scale, however, has been revised and worked over, and from it there have developed a wide variety of instruments for the measurement of intelligence. The application of these instruments to a large number of adults and children has revealed among other things very wide individual differences. What is the social significance of these differences?

One way to answer this question would be to examine the tests themselves and their scoring plans. When this is done we find that the stock intelligence test is composed of a medley of questions, the answers to which require a certain amount of experience but chiefly the ability to manipulate words and figures. The individuals who make the highest scores on intelligence tests are the individuals who are most adept in the manipulation of symbols of communication. Such an ability, however, is a social asset, for there can be no question but that skill in manipulating symbols of communication is a very important factor in determining the position of an adult in his community.

Thorndike (36) has pointed out that there may be at least three rather distinct types of intelligence. The first is abstract intelligence, which is the ability to manipulate abstractions. This is the variety that is measured by the ordinary intelligence test. Second, mechanical intelligence, which is the ability to manipulate tools. Such abilities run all the way from such a simple operation as screwing a nut on a bolt to very complex operations, such as putting together a complicated machine or diagnosing the trouble with a stubborn marine engine. The third variety is social intelligence, which is the ability to manipulate and deal with people. This type might be conceived to run all the way from individuals who simply cannot get along with anyone, at one extreme, to individuals who are able to get along well with everyone and at the same time lead others in all sorts of enterprises. Unfortunately there are no satisfactory tests for this variety of intelligence. One test has been proposed by Moss but it turns out to require the manipulation of abstract symbols and photographs rather than the manipulation of actual persons. What is really needed in this field is some sort of a performance test. Various schemes have been proposed, but none has been definitely worked out and standardized.

A great many studies have been made on the relation between abstract intelligence and an assortment of social variables. The problem has been approached in two ways: first, the persons occupying high and low social positions have been tested and the differences in their intelligence-test scores noted; the other approach has been to study the social backgrounds and the community standing of persons who have high intelligence, on the one hand, and those of low intelligence, on the other. Examples of these studies will now be presented.

The social background of eminent men. Studies of the social background of eminent men reveal that while rise from the ranks is not an uncommon occurrence, yet most of them come from the upper social classes. Havelock Ellis (48) reports that in a study of 829 British men of genius 18.9% came from the nobility, 41.4% from the professional classes, 31.2% from the commercial classes, and 8% from the lower classes. The third edition of Cattell's dictionary of *American Men of Science* shows that of the 885 most eminent (starred) men, 43% came from the professional classes, 21.2% from agricultural, 35.7% from business classes. Professional classes have contributed in proportion to their number 14 times as many scientific men as the other (5 p. 784). Odin reports that of 823 French men of letters, 65% came from the nobility, 23% from the professions, 12% from the commercial and the lower classes. Clarke reports that out of 666 American men of letters 49% came from the professional classes and 7% from the classes of unskilled laborers. Evidently literary and scientific eminence is closely associated with social and economic background. Catherine Cox (12) and others studied the lives of 282 of the most eminent men in history and classified the occupations of their fathers on the Taussig scale. They came out as follows: From the professional classes, 52.5%; semi-professional, 28.7%; skilled, 13.1%; semi-skilled, 3.9%; unskilled, 1.1%; not recorded, 7%.

The social background of the feeble-minded. At the other end of the scale of intelligence are the feeble-minded. In institutions providing for such persons may be found a vast collection of social histories. These histories show that fully 90% or more of mental defectives come from the lower classes. One reason for this, however, is the fact that defective children in the homes of the well-to-do are not, as a rule, sent to an institution. The result is that the institution selects its cases from the lower classes; therefore, the exact proportion of mental defectives among the intellectual and social levels of society is not known, yet, on the other hand, no one doubts that the proportion is greater amongst the lower classes than amongst the upper classes.

Intelligence and occupational levels. Several studies have been made concerning the intellectual abilities of persons who have engaged in different occupations. Many years ago the Harvard economist, Taussig, divided occupations into five levels; professional, semi-professional, skilled, semi-skilled, and unskilled. When the results of the army intelligence tests were classified according to the occupations which the soldiers followed in civil life, there was found a high positive correlation between occupational level as measured by the Taussig scale and intelligence-test score. Fryer (16) took these results, added others to them, and made up a table of ocupations arranged according to the level of intelligence exhibited (Table 1).

TABLE 1

OCCUPATIONAL-INTELLIGENCE STANDARDS—INTELLIGENCE STANDARD INDEX
96 Occupational Designations

Intelligence for achievement			Occupation
Intel. group	Score ave.	Score range	
A	161	110-183	Engineer (Civil and Mechanical)
	152	124-185	Clergyman
	137	103-155	Accountant
B	127	107-164	Physician
	122	97-148	Teacher (Public Schools)
	119	94-139	Chemist
	114	84-139	Draftsman
	111	99-163	Y.M.C.A. Secretary
	110	80-128	Dentist
	109	81-137	Executive (Minor)
C+	103	73-124	Stenographer and Typist
	101	77-127	Bookkeeper
	99	78-126	Nurse
	96	74-121	Clerk (Office)
	91	69-115	Clerk (Railroad)
	86	59-107	Photographer
	85	57-110	Telegrapher and Radio Operator
	83	64-106	Conductor (Railroad)
	82	57-108	Musician (Band)
	81	59-106	Artist (Sign Letterer)
	81	60-106	Clerk (Postal)
	81	57-109	Electrician
	80	62-114	Foreman (Construction)
	80	56-105	Clerk (Stock)
	78	54-102	Clerk (Receiving and Shipping)
	78	61-106	Druggist
	77	59-107	Foreman (Factory)
	75	56-105	Graphotype Operator
C	74	53-91	Engineman (Locomotive)
	72	54-99	Farrier
	70	46-95	Telephone Operator
	70	44-94	Stock Checker
	69	49-93	Carpenter (Ship)
	69	48-94	Handyman (General Mechanic)

	69	46-90	Policeman and Detective
	68	51-97	Auto Assembler
	68	47-89	Engineman (Marine)
	68	42-86	Riveter (Hand)
	67	50-92	Toolmaker
	66	45-92	Auto Engine Mechanic
	66	45-91	Laundryman
	66	49-86	Gunsmith
	66	44-88	Plumber
	66	44-88	Pipefitter
	65	44-91	Lathe Hand (Production)
	65	43-91	Auto Mechanic (General)
	65	43-91	Auto Chauffeur
	65	42-89	Tailor
	65	44-88	Carpenter (Bridge)
	64	43-88	Lineman
	63	40-89	Machinist (General)
	63	46-88	Motor Cyclist
	63	41-86	Brakeman (Railroad)
	62	31-94	Actor (Vaudeville)
	61	40-85	Butcher
	61	44-84	Fireman (Locomotive)
	61	39-82	Blacksmith (General)
	60	38-94	Shop Mechanic (Railroad)
C	60	36-93	Printer
	60	40-84	Carpenter (General)
	59	40-87	Baker
	59	39-83	Mine Drill Runner
	59	38-81	Painter
	58	37-85	Concrete Worker
	58	40-83	Farmer
	58	37-83	Auto Truck Chauffeur
	58	37-82	Bricklayer
	57	41-81	Caterer
	57	39-71	Horse Trainer
	56	38-76	Cobbler
	55	35-81	Engineman (Stationary)
	55	34-78	Barber
	55	35-77	Horse Hostler
	52	38-96	Salesclerk
	52	33-74	Horse Shoer
	51	31-79	Storekeeper (Factory)
	51	26-77	Aeroplane Worker
	51	31-74	Boilermaker
	50	33-75	Rigger
	50	30-72	Teamster
	49	40-71	Miner (General)
	48	21-89	Station Agent (General)
	40	19-67	Hospital Attendant
	40	19-60	Mason
	35	18-62	Lumberman
	35	19-57	Shoemaker
	32	16-59	Sailor
C—	31	20-62	Structural Steel Worker
	31	19-60	Canvas Worker
	30	16-41	Leather Worker
	27	19-63	Fireman (Stationary)
	27	17-57	Cook

	26	18-60	Textile Worker
C—	22	16-46	Sheet Metal Worker
	21	13-47	Laborer (Construction)
D	20	15-51	Fisherman

The column headed "score range" does not mean the entire range of intelligence found within that occupation but the range within which success may be expected within that occupation. This range is usually, but not always, that which is covered by the middle 50%.

While it is true that there is a sharp decline in the average scores of the occupations as we read them down the table from the professional to the unskilled group, yet at the same time there is considerable overlapping; for example, the score range of engineers is from 110 to 183, and this is not the entire range. It is quite likely that the most intelligent plumbers may score as high as some of the least intelligent engineers. Occupational success cannot be predicted from intelligence. The significant point from this discussion is that occupational status, which is indicative of social and economic community status, is correlated with intelligence. This correlation does not tell us whether the occupational level is the cause of the intellectual level, or vice versa. All we know is that persons of high intellectual abilities are more likely to be found occupying high positions in the community.

The intelligence of dependents. Further light on this subject may be secured by noting the intellectual abilities of children and adults who are public charges. It is a well-known fact that children who are found in orphans' homes and in other social institutions are less intelligent than the average run of the population. Pintner (32) has summarized the results of a dozen studies which show that the percentage of feeble-mindedness amongst dependent children varies from 6 to 24, with a median of about 8%. The importance of this figure will be seen by remembering that the percentage of feeble-minded children in the average run of the population outside of institutions is less than 1%. The percentage of feeble-mindedness amongst dependent adults, however, is considerably greater. While there is no study of the general run of dependent adults, there are two or three studies of institutional cases, showing from 20 to 35 per cent of feeble-mindedness. These percentages are somewhat verified by a few tests which have been made on the unemployed, showing that here also the percentage of feeble-mindedness is considerably greater than that in the average run of the population.

The correlation between intelligence and social background. Recently several attempts have been made to secure a quantitative measurement of the socio-economic background of individuals and also of their general cultural levels. A simple test of socio-economic background is that of the Sims score card which attempts to get at the facts about the whole background by asking a few simple questions. This questionnaire has been applied to many thousands of school children and correlations have been found between the results and the intelligence of the children. Hartshorne and May (20) report a correlation of .395 between the Sims scores and intelligence measured by the Thorndike CAVI intelligence test. A test for measuring the general cultural background has been devised by Dr. Edith Burdick. This instrument differs from the Sims score card in many respects. It is built more like an intelligence test, the problems and questions being drawn from home life and from the social and cultural environment of children. The test aims to find out facts concerning the physical conditions of the home but more especially the social and spiritual conditions. The correlations between this test and the Thorndike CAVI intelligence test in the school population in which the home backgrounds vary rather widely and in which intelligence runs from about that of a fourth-grade to that of an eighth-grade child are approximately .50. These correlations are about the same as have been found between the occupational level and intelligence.

Intelligence and citizenship. A great deal more is known about the intelligence of persons who are in institutions—criminals and prisoners—than about people who are citizens. Social failures are always a problem and come in for more than their share of study. Up until very recently it was generally believed that prisoners have a lower average intelligence than the population at large. For example, in 1923 Pintner (32) published a list of studies that had been made, going back as far as 1911, which showed that the percentage of feeble-mindedness amongst delinquent children ran all the way from 7%, in two studies, to 93% in another study. He also reports a dozen studies of intelligence of adult prisoners, showing percentages of feeble-mindedness running from 16 to 54. Pintner, however, points out the fact that when the studies of the intelligence of social offenders are listed chronologically they tend to show decreasing percentages of feeble-mindedness. For example, Hill and Goddard in 1911 published the results of 56 cases of delinquent children, 93% of whom were reported as feeble-minded. In 1922 Healy published the results of tests of 1212 delinquent children, 7% of whom were reported as feeble-minded. Pintner (32) figured that the first 16 reports show a median percentage of feeble-mindedness of 64, whereas the last 16 studies show a median percentage of feeble-mindedness of 26.

The same thing in general is true of adults; the first six reports on the intelligence of adult social offenders show a median percentage of feeble-mindedness of 42, whereas the 7 later studies show a median percentage of 29. Recently, Murchison (28) has published the results of the army intelligence tests as applied to a large number of prisoners in the various prisons of the United States. His results show no significant difference between prisoners and the draft army. The latest information on the intelligence of delinquents comes from Healy and Bronner (21) in 1926 showing that 13½% of their cases were feeble-minded.

The same year (1926), Slawson (33) reported figures on 1543 delinquent boys, among whom 13.4% tested as mental deficients.

While the high percentages of mental deficients amongst delinquents and criminals reported in the earlier studies have been toned down considerably, yet the fact still remains that the incidence of feeble-mindedness or mental defect among the delinquents and criminals is from 5 to 15 times as great as in the population at large. Murchison's study to which we have just referred seems to have been the only exception to this general conclusion. The distinct percentage of defectives amongst social failures is of great importance. Although delinquency and criminality cannot be predicted from a low I.Q., yet one's chances of occupying a position of respect in one's community are considerably reduced by having an intellect that is rated as deficient.

Intelligence and social progress. While social progress is very complex and involves numerous variables, yet it is commonly believed that it is correlated with intelligence. Alexander computed the average Army Alpha intelligence-test scores for the native-born white soldiers of 41 states in the Union. These averages were then correlated with certain measures of social progress. He found a correlation of .79 between the ranking of certain states in intelligence and the ranking of the same states in the number of men whose names appear in *Who's Who in America.* The correlation of the ranking of the states according to the percentage of individuals who own their own homes and the ranking according to the Army Alpha is .68. The states were also ranked on the basis of percentage of literacy (negroes omitted). This ranking correlated with the ranking according to intelligence .64. The states were then ranked according to the educational status as measured by the Ayres index. This gave a correlation with intelligence of .72. Whatever may be the conclusions concerning the causal factors involved, it is certainly true that there is a positive correlation between intelligence and various measurements of social progress. If it is true that the position of any community on the scale of social progress is correlated positively and high with the average intelligence of that community, and if it can be shown that intelligence is a causal factor in social progress, it follows immediately that when the average intelligence is on the increase the community is going forward, and when it is on the decrease the community is going backward.

Certain prophets of doom have come forward and, without stating very clearly the assumptions involved, have seized upon certain results of intelligence testing from which they have forecasted the deterioration and decay of Western civilization. The facts on which these dark forecasts have been made are of the following order. First, the birth rate amongst the lower social classes is increasing more rapidly than that amongst the upper social classes. Since there is a rather high correlation between intelligence and social status, the effect of this differential birth rate is that the I.Q. of our nation is gradually but surely declining. Second, during the first quarter of the 20th century, Europe has been sending us poorer and poorer intellectual stuff in the way of immigrants, until now the average intelligence of immigrants, especially those of southern Europe, is distinctly lower than that of the native-born American. This, again, is tending to pull down the average intelligence of the country. Third, the effect of racial intermarriage is not to raise the intelligence of the inferior race to that of the superior race but rather to pull down the superior to a point between the two. Thus, when the native-born American of "superior intelligence" unites with an immigrant from southern Europe of "inferior intelligence," the result is a family with an average I.Q. lower than that of the superior parent.

The evidence in support of these facts is very voluminous, and for the most part, true. The difficulty is with the implications and assumptions. In the first place, there is no general agreement as to what constitutes social progress. For example, certain aboriginal tribes of Australia, designated by us as backward, may very justly claim a high degree of culture and civilization. While it is true that they do not have telephones, telegraph, trolley-cars, radios, insane asylums, Rotary Clubs, and a Salvation Army, yet they may have achieved a very high order of social values. In the second place, the correlation between intelligence and social progress has not been determined. It may very well be that the significant correlation is between progress and variability or diversity of intelligence rather than the correlation between progress and the average level of the community. In the third place, it will be noted that the argument for doom depends in part on the assumption that intellectual ability is inherited. That heredity plays its part, no one will doubt, but environment also plays a part. If schooling and social environment play the part that some psychologists think they do, then the prophet of doom is left crying into the night.

The foregoing discussion of the social significance of the intellectual ability of an adult may be summarized briefly by calling attention to the fact that whereas intelligence is without doubt one of the greatest of social assets, yet the question arises as to whether it is a cause or a result of social position. Take Cattell's observation that American scientists are 14 times as likely to be found among the professional classes as among the lower classes. Does this mean that the upper classes are likely to breed more brains than the lower classes, or does it mean that they are better equipped in the way of leisure time and stimulating influences to develop what they have? Or, again, does the fact that the accountants, engineers, and college professors score higher in intelligence tests than blacksmiths and taxicab drivers, mean that persons in higher positions are there because of greater intellects, or do they have greater intellects because they happened to be born into the more privileged classes? Such queries are indeed perplexing and raise the whole question of the relative effects of heredity and environment or rather the question as to how heredity and environment combine. This problem has recently been rather methodically investigated by a number of psychologists, especially Freeman (15) of Chicago and Terman (34) of Stanford. The results are too detailed for repetition here. They do show, however, that environment plays a considerable rôle. The study of Hugh Gordon on the intelligence of socially inferior children, as gypsies and canal-boat children, suggests the same conclusion. But whatever may be the final solution to this important problem, there can be no escape from the conclusion that a high I.Q. is a tremendous social asset.

The Social Significance of Individual Differences in Personality. Personality is a complex variable. In fact, it is the most complex of all the human variables because it includes, in a certain way, all of them. It is distinctly a social variable. In this it is like intelligence. What, then, is personality? Many psychologists have written down definitions for it. No two are alike. Psychologists are pretty well agreed that the term personality covers a lot of ground, but they are by no means agreed on the areas that it does not cover. Two or three examples will illustrate this point:

J. B. Watson (38, p. 392): "Let us mean by the term personality an individual's total assets (actual and potential) and liabilities (actual and potential) on the reaction side."

H. C. Warren (37, p. 383): "Personality is the *general rating* of an individual. It embraces all the various phases of one's character—temperament, intellectuality, skill, and morality."

F. H. Allport (1, p. 101): "Personality is the individual's characteristic reactions to social stimuli and the quality of his adaptation to the social features of his environment."

R. S. Woodworth (42): While Woodworth does not formulate a definition of personality in a few words, yet he has a chapter on it and says that it includes four major factors, physique, chemique (temperament), instincts (drives), and intelligence.

It will be noticed that none of these give any idea of what is not included under the term personality. What are its distinguishing qualities? If it includes all human variables, as each of the above definitions indicate, then how is it to be described? A clue may be found in the type of adjectives that are popularly used to describe personality. Different personalities are described as attractive, forceful, strong, magnetic, or their opposites. Practically all of such adjectives refer to the individual as *stimulus*. Thus we may expect to find the differential of personality just at this point. Suppose we set down as a tentative hypothesis that personality is an individual's *social stimulus value*. It is the way in which he impresses others. It includes his physique (all the anatomical variables), his dress, his manners, his voice and the way he handles it, his choice of language, his habitual modes of response, his general and special abilities, in short, all the anatomical, physiological, and psychological variables listed in the table on pages 748-750. All of these combine to make up his total social stimulus value.

This concept differs from that of Watson and Allport in that they find the chief differential of personality on the side of the response. Personality is the sum total of one's socialized responses (Watson), or of one's most characteristic social responses (Allport). The notion seems to be something like this: that if an individual were measured in all the possible human variables, on a series of scales yielding comparable scores, his personality would be his total score; just as his intelligence is regarded as his total score on a battery of intelligence tests. In making up this total personality score different variables would no doubt have different weight.

for example, social intelligence (the ability to manipulate persons) would have a higher weighting than, say, reaction-time. While this view of the matter is quite legitimate, and while it is convenient to have a name for the total score, yet this is not what the layman regards as personality. Personality, to the man on the street or in the office, is a concrete reality. But personality, as the psychologists view it, is somewhat of an abstraction. When Watson says that it is one's total assets and liabilities on the reaction side, and goes on to define assets and liabilities as socialized reactions, he has in mind, I judge, a kind of algebraic summation, or average; while Allport seems to have in mind something like a mode in a frequency curve when he talks about the most characteristic ways of social response.

The view presented here, which finds the differential of personality on the side of the stimulus, avoids many difficulties. In the first place, it is in line with the everyday, common-sense notion of what personality is. In the second place, it makes it concrete and real. And in the third place, it provides an avenue of approach to the scientific study of it.

As stimulus value, personality ranges all the way from zero to some unknown amount. The individual who has zero personality (if there be such a person) is one who makes no impression on anyone. As social stimulus his value is zero. At the other end of the scale is the individual who as stimulus produces in others very intense and vivid responses. Thus the measure of a personality is found by counting and measuring the intensity of the responses produced in those with whom it comes in contact. Psychologists have long since recognized that responses vary in intensity from weak to strong, also that their affective accompaniments (perhaps other responses) vary as well. This effective variation is known as the law of affect. Now suppose we wish to measure Mr. A.'s personality—we should do it by measuring the strength and affective accompaniments of the responses that he produces in all others with whom he comes in contact. He will affect different persons differently, but his sum total is his personality score. Or he may have a different score for each person with whom he comes in contact, and his stimulus value for the same respondent may not always be exactly the same.

One of the favorite indoor sports of certain psychologists is to analyze personality into traits, or component parts. Most of the outstanding personalities are supposed to possess certain traits in rather marked degrees. It is therefore believed that the way to understand personality is to analyze it into traits. Now what is a trait? Well, it is any variable, such as height, weight, cephalic index, memory, intelligence, honesty, etc. Traits, too, are usually viewed as response values. But they may also have stimulus value. As such they are the different ways of impressing others. For example, Mr. A. may have a very smooth, musical, and appealing voice. When those who have heard him speak are asked to describe his voice (and what they describe is the way they respond to the stimulus of his voice), they do so by saying that it is smooth, musical, pleasing. If his voice impresses others as muscial, then he has a musical voice. Again, if Mr. A. gives others the impression of refusing to associate with them, of assuming a supercilious air of superiority, he is said to be snobbish. But his snobbishness is not something that he possesses, but something inferred by the way he carries his cane, or wears his necktie, or speaks to others. Others react in a more or less unified and similar manner to certain of his responses or attitudes, and these unified responses are described as traits that he *possesses.* As a stimulus, he touches off typical reaction patterns in his associates. It is these patterns that are called traits. Thus his traits are not something that he possesses, but rather, constant stimulus values.

Personality traits vary enormously in their social significance. The Allports have singled out a trait which they call "ascendance-submission" as one of the most important. In this connection it will be remembered that McDougall (27) bases his whole theory of morality on the instinct of self-assertion and submission. But the Allports are not after instincts. Ascendance and submission are viewed as typical social response patterns. They describe the behavior characteristic of each. Yet the table might be turned and permit them to be viewed as stimulus values. The ascendant individual is a person to whom others respond by bowing down; the submissive person is the one to whom others respond by dominating him. In fact, such traits seem to fit the picture better as stimuli than as responses.

Now the point to all this is that the place that any individual occupies in the community is partly, if not mainly, his social stimulus value. He may be a weak or strong social stimulus. If he is a strong social stimulus he will occupy a prominent part; if he is a weak social stimulus he will play a minor rôle. But since social participation has not been measured quantitatively (although many interesting attempts are being made) and since personality has not been measured quantitatively, there is no way to know just what the correlation between the two is. Some day when these measuring instruments are more perfected, we may know. In advance of the fact, however, we may well guess that the correlation is high but not perfect. For the community leaders are not always the strongest personalities. Leaders in world affairs are, however, as a rule, men of strong personalities. For example, the Big Four of the Peace Conference, Wilson, Lloyd George, Clemenceau, and Orlando, were all strong personalities.

The measurement of personality today is about as far advanced as the measurement of intelligence was 25 years ago. We are still in the stage of attempting to measure traits with the hope of combining them into total personality profiles or scores. It is doubtful whether we shall ever have a personality measuring scale that will yield a single total score analogous to the I.Q. or point score on a group intelligence test.

On the response side, the characteristic thing about personality is integration. The problem is to determine how all the human variables are integrated into one working machine, and many such machines are integrated into a larger social community. It is to this problem that the last section of this chapter will be addressed.

INTEGRAL SOCIAL PSYCHOLOGY

Properly speaking, this chapter should begin with a definition of the terms "adult" and "community." For present purposes the dictionary definition will suffice. "An adult," says the dictionary, "is a person, animal or plant grown to full size and strength; one who has reached maturity." A community is "a body of people having common organization or interests, or living in the same place under the same laws and regulations . . . An assemblage of plants or animals living in a common home, under similar conditions of environment, or with some apparent association of interests." A mere assemblage or aggregate of individuals is not a community; before they can become a community they must have something in common. The common factor that makes them a community must be integrative and dynamic. They might, for example, all have red hair or blue eyes and still not be a community. The common factor or factors are not structural but dynamic; they may be structural, too, but this is not important. In short, a community is an aggregate of individuals that "acts as a whole," that has in it a certain order and unity of a dynamic sort. The problem before us is to determine how a community or group can act as a whole.

The point of view adopted in this discussion of integration is frankly biological. It is recognized as one of many points of view that could be taken. But of all the possibilities this seems to be the most promising, mainly because it is supported by the most factual evidence. The support, however, is mostly indirect, since the biological theory of social organization is mainly derived by that rather dangerous proceeding known as reasoning by analogy. But the analogy is not necessarily the dangerous one that is hung up as one of the red lanterns in all textbooks on logic, for the relation between social organization and biological organization is more than analogy. It is based fundamentally on an *assumption of continuity*. Since all theories start with some assumptions, stated or otherwise, one is as good as another; in fact this one, biological continuity, seems better than the average, if for no other reason than that it stimulates research.

The plan of attack is first to find out if possible the mechanisms and processes of integration in lower biological organisms, then in higher animals, then in animal communities, and finally in human communities. Can we find, therefore, a single principle of integration or theory of integration that will explain all of these? To do this would necessitate bringing together all the facts known about integration at all levels and see if they can be subsumed or accounted for on a single hypothesis; or rather, see if they do not have points in common, or so nearly in common that they may be regarded as manisfestations of a single principle of integration.

Integration in Lower Organisms. In this section we shall rely mainly on the data collected by Child (7), who seems to have given long and careful consideration to the problem and has come out with a definite and apparently well-supported theory. At any rate, his analysis of the facts of integration in lower and higher organisms leads immediately to a social theory that is not unlike that of Herbert Spencer, which, as is well known, has had a tremendous influence even though it is no longer generally accepted.

Child sets out to explain the origin and nature of the observed order and unity of behavior in organisms. It is assumed that, since all organisms do exhibit orderly and apparently unified behavior, there is some principle back of it which, when known, will account for the observed facts. He begins with the notion that an organism is defined by its relation to the environment, and organismic behavior is that of the entire organism. Structurally an organism is composed of specific protoplasmic parts, each of which has its own behavior pattern. Protoplasms are highly specific, not only in the same species but even in the same individual. The changes in the protoplasmic patterns, or protoplasmic activity, is the simplest form of life, and the whole organism is the integration of these protoplasmic integrals.

Now an organism has two kinds of relations to its environment, material or transportive, and dynamic or transmissive. In the former, substance or energy is transported from the environment to the organism (import) and from the organism to its environment (export). The dynamic relations are transmissive in which energy is transmitted back and forth between the organism and its environment. This is done in two ways: first, by purely mechanical means in which the effect is always physically proportional to its cause, and, second, in an excitatory way in which the environment seems to liberate stored energy within the organism, and the cause is not apparently proportional to its effect. Child makes a great deal of this excitatory kind of transmission and thinks it is the most important of all. By it any part may assume control over other parts without direct mechanical contacts with them. It makes possible the mechanisms of dominance and submission which describe how it is that one part may at any time be dominant over others.

In order for an organism to function as a whole and in a unified way, at least three things are necessary: (1) a differentiation of parts; (2) communication between these parts; and (3) some sort of control. Each of these requirements is met in lower organisms, at least, by the principle of "physiological gradients." This is the name given to the observed fact that in any organism at any given time different rates of living may be observed. By different rates of living is meant differences in metabolisms of various sorts, differences in chemical and physical velocities, differences in colloid changes, and others. Such differences, however, appear to be scaled or graded since there is always an area of maximum activity which shades off into areas of minimum activity. Child marshals a considerable body of evidence and puts up a convincing case that such graded differences in rate of living do in fact exist, and, what is of much greater importance, can be controlled within limits. The scientific data upon which he bases his conclusion that such differences do exist in all living organisms cannot be included here. The inquisitive reader will find them in his volume on *The Physiological Foundations of Behavior* (7).

We have just noted that the "physiological gradient" is for Child the physiological mechanism of integration, *par excellence.* In the first place he uses it to explain physiological differentiation. By controlling the gradient he can, within the limits of hereditary potentialities, produce at will almost any kind of parts. For example, the most characteristic structural pattern of lower animals is the "axiate" pattern with its longitudinal polarity and its latitudinal symmetry. By altering the gradients in various ways polarity and symmetry may be altered, such as developing individuals with small heads, with one eye in the median plane like the mythical Cyclops. By obliterating the gradient entirely he is able to obliterate completely polarity and symmetry in such a way that the organism remains spherical with no differentiation except that of interior-exterior. Also by introducing new gradients he can experimentally produce new structures. In fact, by altering the physiological gradients at very early stages of development he is able to produce individuals that differ more from their own species than species differ from one another. From these experiments and observations he is convinced that differentiation among the parts of an organism is a direct result of the differential living rates of the organism.

The second requirement for integration, that of physiological communication, is explained by the same principle. In the first place, it is noted that the most active region is always the dominant region, and the less active, the subordinate region. The most active region is the region of the peak of the gradient from which the activity is graded down to the less active regions. Of course, the activity of the dominant region must be communicated to other dominated regions. This is done by the two processes of communication mentioned above, namely, transportive and transmissive.

This at once furnishes a ready explanation for the principle of physical control; for given (1) the principle that the most active dominates the less active, and (2) that the activity may be readily communicated, we have all that is needed for control. Furthermore, Child finds that certain gradients tend to be persistent and thus form a basis for learning. The facts supporting the principle of dominance and subordination are numerous. In animals without any nervous system it is easy to demonstrate experimentally that the most active areas are always in command. In organisms that have a head and a nervous system it is a matter of common knowledge that the brain is the most active organ (at least it consumes the most oxygen). The central nervous system is in his estimation the best structural example we have of a gradient.

Integration in Higher Organisms. When higher organisms are dealt with, Child admits that experimental control is more difficult. Apparently the gradients are harder to control and the hereditary patterns are more set. Yet on the theory of physiological continuity, he maintains that the principle of integration that so neatly explains the "order and unity" in the behavior of lower organisms will also explain it in higher organisms and in social organizations. (In a later section we shall return to his theory of social integration.)

The principle of physiological gradients as an explanatory hypothesis for the integration of higher organisms is not without some basis in fact. In a treatise on the "Origin and Development of the Nervous System," Child brings together various lines of evidence which indicate that the structure and functions of the nervous system originated in excitation-transmission gradients. The evidence is too extended to be presented here and indeed it need not be. It is necessary only to point out that the nervous system is the great system of communication between the parts of higher organisms, and that the brain is the dominant organ. There is no denying the fact that higher organisms possess all the elementary requisites of integration, namely, differential parts, communication between them (both transportive and transmissive) and dominance and subordination (i.e., control). But whether these processes can be explained by the principle of "physiological gradients" is quite another matter.

No story of the integration of higher organisms would be complete without reference to the great work of Sherrington. It is impossible within the brief compass of this chapter to summarize it and point out its significance. Among the most important contributions of Sherrington are the discoveries of the anatomical processes of convergence and divergence in nerve structure; of the functional processes of facilitation and inhibition; of how it is that a single sensory stimulus may excite different motor responses; how the same motor response may be made to many different sensory stimuli; of how reflexes are timed and how they work in teams; of how in completing any movement or act of skill the muscles work in teams with the proper time relations beween relaxation and contraction; how certain reflexes have the right of way over others, and how this right of way or prepotency is determined; and of how postural reflexes make ready and throw into gear whole chains and systems of reflexes which work in harmony together toward a common goal. These and other important facts lie back of a complete understanding of how the individual functions as a whole.

Allied with such studies as that of Sherrington on the *Integrative Action of the Nervous System* are studies of the integration of other bodily organs, especially the glands and the smooth muscles which are not connected directly with the cerebrospinal system. Investigations of the emotions have revealed many facts showing that in an emotional crisis, and even in minor emotional states, the organism functions as a whole. Consider the case of fear. When the animal perceives that he is in danger, his whole organism prepares either for flight or for fight. The stomach stops digesting, the heart beats faster, the breathing is accelerated, certain glands go on extra duty and pour out their substances into the blood stream, still other glands shut down operations entirely. All this goes on in perfect harmony in spite of the fact that the body appears to be in great upheaval and in great perturbation.

The work of Sherrington has been utilized by investigators of the learning process, especially by those studying the ways in which motor skills are acquired. Such skilled performances as playing a piano, running a typewriter, sailing a boat, playing tennis, represent a high degree of integration. Any one of these requires an elaborate process of inhibitions and facilitations, of timing, of determining the relative strength and weakness of the component responses. Detailed studies of how these elaborate series of responses can work together in teams for the achievement of the varying degrees of success that attend the performance have revealed certain mechanisms which are now understood. But no one as yet has been able to describe in detail all that happens, much less offer a satisfactory explanation of the intricate process. One such process of integration that has been observed is that of "higher units" (Woodworth, 42). The best illustration is learning to typewrite or telegraph. First the learners established certain "letter habits," then with more practice and more "learning" they established a higher order of "word habits," then "phrase habits," and finally "sentence habits." The basic concept here is that of hierarchy of habits. Another integrative mechanism is found in the principle of the conditioned reflex which has been discussed at length elsewhere in this volume. This principle has been generalized by Hollingworth (22) into a more inclusive one which he calls "redintegration." Hollingworth finds this to be quite general and to furnish a basic explanation for mental life as a whole. The principle is briefly this: Any part or relation between parts of a total situation may come to stand for the whole, so that the organism responds to the part in the same way it does to the whole. In this way a very complex stimulus may be compressed or represented by a very small feature of it. This stimulus acts as a whole by representation, just as a community delegates power to a representative of it to act for it at a convention.

Another allied line of investigation of the processes of integration has been pursued by the Gestalt psychologists with their insistence on the fact that the whole is always greater than the sum of its parts. Their work has been mainly in the field of visual perception and the major problem is that of understanding the perception of patterns. A very simple illustration is that of any geometric arrangement of lines or dots, e.g., this :: Here we have four dots, physically nothing more, but psychologically a square. Whence the square? It is not contained in dots taken separately, but in their relations, which are perceived as a total Gestalt or pattern. Likewise with a melody, which is something more than the notes or sounds that compose it. The sounds and their relations form a pattern which is perceived or reacted to as a whole. The Gestalt psychologists contend that these patterns (configurations) are the irreducible datum of psychology. They are integrations, wholes, and behave as such. Likewise, the individual as a whole is a pattern, a configuration of habits.

Individual integration has been studied from still another point of view by the psychiatrists, psychoanalysts, and abnormal psychologists. The question is approached from the side of disintegration. Such phenomena as mental conflicts, emotional conflicts, and ultimately divisions of the personality, have been studied in great detail. The interest here is to work out practical programs to prevent or to teach the individual how to prevent disintegration. A great deal of mental hygiene is just this. The idea is that a well-integrated mind is a healthy one. While the literature is vast and vague, it is enough to say that these treatments or preventions usually take the form of finding the source of the conflict and uprooting it at its base. The theory of integration as found in this literature is in the main a negative one—the avoidance of conflict. Since the subject of conflict will perhaps receive its share of space in another chapter, it may be dismissed here.

All of these endeavors to understand adult integration seem to end in one general conclusion. It is that integration requires as its minimum essentials the *communication between parts* and the *dominance* (temporary at least) *of one part over the others.* Another conclusion that can be drawn is that integration can best be understood from the genetic or developmental point of view. No cross-section will give a complete picture of how it happens. It is the result of a long history of development of parts, and of functional communication and control between them, due in the main to the differential effects of the environment working within certain limits set by heredity.

Integration in Animal Societies. Considerable scientific work has been done in the field of animal sociology. The social life of various species of animal life has been studied in detail, with careful descriptions of social activities. By social activities is meant activities observed among animals when they work together. The most noticeable thing about these detailed descriptions of animal life is the wide variety of activities involved and the almost insuperable difficulties of classification. In fact the impossibility of categorical classifications is one of the outstanding features. There are no definite categories, just as there are none for human behavior. However, many of these descriptions may be summarized by noting a series of variables:

1) The first and most noticeable variable in animal behavior is in respect to sociability. They range all the way from the solitary animal, that spends most of its life alone, to the most complex social groups in which the members spend most of their lives in the company of their kind. It would be possible to make a scale on which could be placed various animal societies. Toward the solitary end of such a scale would be those animals that come together only at mating time; a little higher up would be a place for those that live together long enough to rear the young, after which the family separates, and so on up for the more and more complex societies.

2) Animal societies vary in respect to requirement foı membership. At one end of such a scale are the "wide open" societies which may be joined by any other animal, even of another species. Higher up are societies that will admit members of their own species only. Still higher on the scale the entrance requirements are more and more rigid, until the upper end of the scale is reached, which would represent the "closed tight" order that admits no one.

3) Another variable is that of degree of leadership. At one end of

the scale there are groups which have a definite and permanent leader or boss; at the other end there is no leadership, or else it is a matter of chance. An example of chance leadership is noted in some species of fish travelling in schools or shoals and following any member of the group who happens to be on the periphery. At the other end of the scale definite leadership is found in cases of sheep that always follow the bell wether, or of cows following the bell cow, or of wolves following the gang leader.

4) Another variable is that of differentiation of functions. Such a scale runs from instances in which there is no division of labor, each member doing the same thing, or nearly so, up to the very highly specialized functions found among ants, wasps, and termites. As is well known, the division of labor in certain species of ants is very elaborate. Some do the housekeeping, clean the nest, and tend the young, others go foraging, others tend the fungus gardens, others keep the "aphid" cows, others look after the "guests," both welcome and unwelcome, and others are guards or sentinels.

5) Animal societies vary also in respect to property rights. Some species find a nest or home and occupy it long enough to rear one brood of young, and then abandon it and pass on. Other species "squat" on one locality, be it tree, or stump, or hill, and stay there for generation after generation, until either exterminated or driven away. For example, each ant state has its own hunting ground, which its soldiers guard against all invasions. It is seldom that property rights of animals extend much beyond the nesting or homing or hunting ground. Most other types of property have to do with such things as food which has been stored up for the winter.

6) Animal societies also vary in respect to the complexities of communication. At the lower end of the scale only crude imitation is noted. That is, one animal in the group responds to the others or to another by imitation. Imitation has been played up considerably in animal and human psychology, but it appears to represent a rather low level of social stimulus and response. Higher up the scale the animal responds to signs or cues, as flying away or running away in response to a warning cry. Still higher up we find instances of animals following "leads." At the very top we have the typical human response to abstract symbols.

The question of animal languages is still in an unsettled state. All sorts of modes of communication have been observed. Some animals communicate with each other by making sounds, that is, by auditory cues. Others communicate by touch, as the supposed antennae language of ants and bees. But it is rare that visual cues are important, even communication by smell is more common, especially in insect societies. Among ants, termites, and bees, communication is carried on by sound, touch, and smell, rarely by sight. Just how systematic and symbolic these signals are, and just how much the behavior of one member is influenced by the actions, sounds, and odors of others, we do not know.

It would no doubt be possible to go on and discover other ways in which animal societies vary, but the above six will suffice to illustrate (1) that the facts of animal behavior cannot be neatly pigeon-holed into clean-cut categories, and (2) that the more integrated societies tend toward one end of each scale. If a large number of species of animals were carefully studied and then rated on these and other scales, it is quite likely that there would be high correlations between different scale values. For example, those societies in which the members are together constantly and all their lives are also most likely to be the very societies to which the entrance requirements are most rigid, in which leadership is most definite, in which property or homing is most guarded, and in which communication is most complex. If this is true, and if these scales were arranged so that the high score would always be toward the more complex end of the scale, then the intercorrelations would probably be high and positive. While quantitative treatments of this wealth of descriptive data have never been made (at least not to the knowledge of the writer), yet an examination of it indicates that such correlations would be high. All of this would mean that animal societies could be arranged on a scale of integration, which, to be sure, would be a multidimensional scale like intelligence-test scales but would nevertheless be valuable. A society high on the scale of integration would be the *Atta* ants, and one low would be a school of fishes or a flock of birds.

7) There is one other variable, however, which may, after all, offer the key to the whole situation, and that is modifiability or plasticity. Earlier in the chapter we called attention to the fact in discussing the matter of how animals adapt their habits of life to their environments. Wheeler (39) seemed very much impressed with the differences in plasticity between the bees and ants. The bees seem forever doomed to build their nest on the same pattern, whereas ants vary the pattern to suit the conditions. There seems to be a high correlation beween adaptability to environmental conditions and complexity of the society and its integration. It would appear that in some species of animals the hereditary limits are set closer, leaving less room for environmental adaptation than in other species. The solitary spider builds its complicated nest without blue-print and without previous instruction, as do certain species of wasps. The process is surely unlearned, yet mechanically true to type, varying only in minor details.

But of the ants, Wheeler says: "Very long and intimate contact with the soil has made the ants singularly plastic in their nesting habits. While most social wasps and bees construct elaborate combs with very regular, hexagonal cells of such expensive substances as paper and wax, the ants merely make more or less irregular galleries or chambers in the soil or dead wood, or if they construct paper or silken nests, avoid a rigid type of architecture. Hence the great variability of nesting habits in the same species. This plasticity and saving of time and labor are very advantageous, because they enable the insects, when conditions of temperature or moisture become unfavorable, or when bothersome enemies settle too near the nest, to change their habitation readily without serious loss to the colony" (39, p. 153). Again, "The plasticity of ants is shown even more clearly in their care of their young, which are not reared in separate cells but in clusters, and lie freely in the chambers and galleries of the nest where they can be moved about and easily carried away or hidden when the colony is disturbed or the moisture and temperature conditions are unfavorable . . . A similar plasticity characterizes their feeding habits. As a group they feed on an extraordinary range of substances; the bodies and secretions of other insects, seeds, delicate fungi, nectar, the saccharine excreta of plant-lice, scale insects, etc. Some species seem to be almost carnivorous."

This plasticity and modifiability go hand in hand with both differentiation and integration. But association does not prove causation. However, if plasticity is defined as multiplicity of hereditary possibilities, it would seem to precede both differentiation and integration. Wherever the hereditary possibilities are the greatest, the differential effects of environmental influences are greatest; and wherever this is greatest, there is the greatest differentiation and integration. Thus our facts seem to hang together. The animal group functions as a whole in the same way that the individual organism does. When the hereditary patterns in the individual organism are wide, then environment has the greatest differential effect and there is the maximum opportunity for gradient integration. The same is true of communities. If the problem is approached from the solitary behavior of an individual or from the social behavior, the explanation is the same.

Integration in Human Communities

The problem. The problem is to find out how it is that human beings work together in teams more or less organized for a common good or goal. The problem is not, as some have supposed, to explain why man is a gregarious rather than a solitary animal, or why it is that human societies exist at all. The existence of societies and communities is taken for granted, assumed as a starting-point, and the problem is to understand the relation between the individuals composing them. What sort of relations are most conducive to the welfare of the community, i.e., to the group? For, after all, the group is the unit of society, and not the solitary individual. How does it happen that the joint efforts of ten men working together will produce greater results than the efforts taken individually and then added arithmetically? What are the maximum limits of such joint efforts—how much more can ten men accomplish when working together than they can singly? What is responsible for the added increment? What types of individuals work best together, or what types form the best integrations? These questions are at the bottom of the integration problem. They are but a few of the questions that the social sciences are trying to answer. They will serve to illustrate the fact that underlying all of them are the basic problems of integration, which are problems of differentiation, communication, control, and motive power.

Human communities vary, of course, in many respects. We have just noted seven variables that may be found among animal communities. Human communities have all of these and more. Perhaps the most important variable that could be added to the seven noted above is that of the degree of commonness of purpose or objective. Even the same community will vary in this respect from time to time. For example during the war the nations composing the Allies on the one hand and the middle European countries on the other were more closely knit than ever before. It requires a common danger, a common goal, to unite people most intimately. Now all communities vary in respect to the degree in which members have a common interest or objective. At one end of the scale there are communities in which the common tie does not reach beyond the mutual protection from danger; at the other extreme, there are communities in which there are a great many common interests and purposes, such as good roads, artistic surroundings, health and sanitation measures.

The concept of integration as applied to communities may be broadened by extending the scale in a negative direction. The antithesis of integration is disintegration. At the extreme negative end of the scale there are communities in which there is a maximum of social conflict, in which every man's arm is raised against every other, in which not only is every man for himself, but everyone is *against* every other one. Such communities have at times existed but not very long; they are soon self-destructive. Moving from this extreme negative position on the disintegration-integration scale to the middle or zero position, we find there a community in which every one is for himself but not against others, i.e., the members act so as not to interfere with the rights of others. At the extreme positive end all pull together for the common good and each acts so as to help the other.[1]

We have already noted that the four major elements of integration, or the four major principles involved, are differentiation, motive force, communication, and control. Differentiation was discussed in the first part of the chapter. We shall now consider the other three.

1) *Social dynamics.* Before we can understand how an organism functions as a whole, or how a community functions as a whole, it is first necessary to understand how it functions at all. Social behavior is some function of the behaviors of the members of the group.

The problem before us is to understand how it happens that individuals get together, play together, work together, live together, and die together. Most plants and some animals are entirely solitary and spend their entire lives, or most of them, living apart from others; but all humans are social in the sense that they live most of their lives in company with others. This is the great fact of social psychology. How did it happen?

Wise and sagacious men have considered this question for many years and several answers have been given. Some of them will be listed here.

a) The most common answer is that man has inherited social instincts of gregariousness, imitation, and the like. These instincts are strong drives, so strong that their driving forces cannot be diverted—they must be satisfied. So the answer is easy. Men live together to satisfy their instincts. But this answer is so easy and so simple that it at once arouses suspicion. Whence came these strong instincts? Ah, there's the rub! The answer is that they have developed out of long years of social experimentation. The philosopher Hobbes had it all figured out this way. The *homo sapiens* was originally a solitary beast, but he took to living with other beasts because he feared them. The first societies were a kind of mutual protective associations, not only against enemies of nature and of other species but against their own groups. In short, men originally lived in groups to keep from eating each other up. Bentham later developed the idea that the first societies were mutual aid societies in which the motive was, "You help me and I'll help you." From this sprang the famous social contract theory which has it that a lot of solitary animals got together and entered into mutual contracts in which every individual used every other as a means to satisfying his own desires. Tarde, Baldwin, and others developed the theory of imitation. When sheared of its exalted language, this theory comes down to this: one beast observes the behavior of another and imitates it, after that they are brothers. A third one comes along and imitates them; that starts a custom or a jungle fad. The contagion spreads and a group is formed. Then beliefs and superstitions arise and are standardized by imitation; purposes or objectives are set up and stereotyped in the same way. In some such fashion as this, primitive tribes are established and nations are born.

---

[1] This is essentially the same as Herbert Spencer's idea of moral levels.

Other similar theories have been developed, but always on the same pattern. The reasoning back of all of them is that men live together to satisfy certain instincts which have developed in curious ways. They all assume that man was first a solitary animal and got to be social at a later date. There are many variations of the instinct theory. A variation of Hobbes's fear theory is the "comfort" theory, which is that men are just more comfortable when in the presence of others or when they are in groups.

The best known of all the instinct theories is that of McDougall (27), who makes use of a wide variety of instincts—seven in all. By taking in a wider range of instincts, he is able to explain more social phenomena than he would by limiting his theory to one or two. Yet McDougall cannot resist the temptation to play up some instincts more than others, so he picks out self-assertion and submission as trumps. He, too, begins with man enjoying his instincts. But by and by man runs up against others of his species, and conflicts arise, in which he either wins or loses. Next time he is more careful. He develops a regard for his fellows. He subdues some and is subdued by others; in this way the self-assertive and submissive instincts get going. He becomes appreciative of praise and blame and has high regard for his superiors. McDougall, however, is of the opinion that the earliest social groups were families, which view is shared by many others. Families developed by virtue of the necessity of caring for the young and propagating the species.

b) In recent years the whole instinct theory has been under severe fire. Instincts are not a part of the modern psychologist's stock in trade; they are out of style. When the psychologist let out the instincts by the back door, he took in substitutes by the front door. If social behavior cannot be explained by instincts, that is, if men do not live together to satisfy their instincts, then why do they live together at all? The simple answer is—because they always have; they can't do anything else. At once we have given away the trick. Man was never a solitary beast, after all. Ever since he became a man he has lived in some sort of group, and there has always been social behavior. To explain how it happened would be to explain the difference between the solitary wasp and the social ants.

Thus Woodworth (41) has recently advanced the hypothesis that the driving force behind social behavior is the same as behind any kind of behavior, namely, the mechanism itself. Man enjoys social activity for the same reason that he enjoys music, or mathematics, that is, because he has the *capacity for it.* This capacity he has always had in varying degrees ever since he became a man. He points out that the social drive is not so much just being in the presence of each other, but in joint activity, in doing things together. The social motive is the tendency to engage in group activity. This group action is interesting and satisfying because of the capacity for it or the ability to do it. But whence this capacity? To quote Woodworth again: "The social motive—and this is the main contention in this whole discussion—is inherent in social activity. Possessing, as he eminently does, the capacity for group activity, man is interested in such activity. He needs no ulterior motive to attract him to it. It is play for him. His interest in it comes partly from the interplay of personality (which he has a native capacity to apprehend), partly from the coordination of the acts of several performers into one harmonious and well-directed action, partly from the spirit of rivalry that may be engendered between groups, and not least from the big enterprises that can be carried through by joint action. In short, the social interest is part and parcel of the general *objective* interest of man" (41, p. 202).

It is part of the general hereditary behavior pattern of the individual which may or may not be realized. Social activities may be initiated in various ways, but once they get going they develop, like everything else, their own drives. Social skills are learned just as other skills are. There is much experimental and observational data to bear this out. Cason has shown that gregariousness is really after all only a common habit, and Anderson (8) and others have shown how social activities are the earliest activities of young children. A child has more opportunities to learn to be social than anything else. These opportunities are so abundant that formal social schooling has up until recently appeared unnecessary.

2) *Communication.* The second important factor in social integration is communication. Its importance may be seen by considering for a moment the integrative action of a machine. An automobile functions as a whole partly because there is communication between the parts. This communication is of two different sorts. The first is mechanical. This is illustrated by such mechanical devices as a clutch and a gear, by means of which the power of the engine is transmitted or communicated to the rear axle of the machine. The second type is transmissive, and is illus-

trated by the electrical equipment. Here the energy of the battery is transmitted to the spark plugs. These two types of communication are found in all living organisms. For example, the action of the heart is communicated mechanically to the veins and arteries in all parts of the body. The nervous system, on the other hand, illustrates the transmissive type of communication by which energy impinging upon sense-organs is transmitted to muscles. It has already been pointed out that the integrative action of the organism, especially of higher organisms, is dependent upon this type of communication.

The integration of a community is dependent upon and is in a degree proportional to its facilities of communication. The chief instrument of communication is language. The story of the development of language is indeed a long one and cannot be told here. It should be pointed out, however, that one of the chief obstacles to international integration is the fact that different languages are spoken by a great many different nations of the earth. Language is the medium by means of which one organism communicates with another. Here we find the basic reason for regarding a community as a biological organism. One individual communicates with another either by the transportive or mechanical method, or by the transmissive method. Within any one organism the different parts function as a whole because they are tied up with a vast system of communication. This system within the organism is not only analogous to, but is essentially the same as the system of communication between organisms.

A single example of this type of communication should make the point clear. As I sit at my desk, I feel a draft of cold air. It stimulates certain areas of the skin. These stimulated areas become excited and at once dominant. I react by rising and closing the window, thereby restoring equilibrium and satisfying the urge or completing the configuration. On the other hand, I may ask some one to close the window. In this case I respond by saying to another person, "Please close the window." In so doing I communicate energy transmissively to the sense-organs of another organism, which causes it to function exactly as I would have functioned had I arisen and closed the window myself. The important factor here is that of communication between organisms, which is exactly the same as the communication between the parts of a single organism. There are all sorts of means of communication between individuals, but always through physical media, just as there is communication between the different parts of a single organism. In commenting upon this point, Child observes:

"Sociologists have often pointed out that the tool, the instrument, in fact, any fabrication, becomes in a sense an organ when used by man in the control of the environment. We may go a step farther than this and say that speech makes the atmosphere, sight the ether, an organ of social correlation, as truly as protoplasm is an organ of physiological transmission. The importance of this concept is great when we consider that the physiological basis of all forms of intercourse, barter and exchange is based on dynamic correlation" (7, p. 268).

So we have at one extreme the simple, unicellular animal which functions as a whole in its confinement by virtue of the dynamic functions of its parts, and, at the other extreme, the complex, social group which functions as a whole in precisely the same way, namely, by the dynamic communication between its parts. It is a common observation that improved methods of transportation and communication, such as aviation and radio, are factors of tremendous importance in the integration of society. National understandings and ultimately world peace are being largely facilitated by these improved methods of intercommunication.

3) *Social Control.* The third important factor in social integration is control. The automobile will again serve as an illustration. The function of the automobile is transportation. It serves this function only when under the control of the driver. When any engine functions in a unified manner, there is always a controlling mechanism. In the individual organism the controlling mechanism, according to Child, is the physiological gradient. The most active part is the dominant part; it is in control because it is sending out the energy. Child has assembled an abundance of evidence to show that the brain, for instance, is the controlling mechanism because it is the most active organ of the entire body. For example, it has been known for a long time that the brain consumes more oxygen than any other organ of the body.

No social integration is possible without some sort of dominance, leadership, or control. In gang life there is always a leader who is the "master mind," or the brains of the group. In certain social groups this master mind or dictator or power of control may not necessarily reside in one organism. In the lower forms of life it often does, as, for example, the leader of the wolf pack or the buck of the buffalo herd, and so on. On the other hand, there may be a committee or a congress, but even here there is usually a boss politician or some one whose leadership is recognized.

The history of social leadership shows varying powers among leaders. The leader is not always the strongest man; he may be the oldest or the most cunning, or the medicine man or the magician. In all cases he is an extreme deviation on some variable, such as strength, cunning, superstition, or reverence. In the higher forms of social integration, the dominance often goes out of the hands of a single man and is crystallized into law, customs, traditions, and social sanctions.

In most social organizations there is limitation to the powers of the dominant person, idea, custom, or force. These limits are set by various conditions. One type of condition is that of ease of communication. In order for the leader or for the idea or for the custom to dominate the behavior of a group, it must be communicated to all members.

Another limitation is the degree of diversity or variance within the group in respect to any of the variables in which its members differ. Social integration, therefore, is dependent on the proper degree of homogeneity in the group. An illustration of this may be found in the matter of color. When any group of individuals is composed of individuals who vary in color from very black through the yellow and different shades of tan to the very white, social integration is much more difficult than it would be if all the members of the group were black, brown, or white. Variations in individual ability and other psychological variables produce other difficulties for integration.

This leads to a consideration of, and offers a hypothesis for the explanation of social disintegration. The integrated activities of communities vary all the way from chaotic social conflicts, on the one hand, to the absence of conflict, on the other. In short, it may be said that the integration of any community is measured inversely by the degree of conflict found therein.

Looking to the future, it may be said that the major problem of social psychology is to discover and chart the areas of social conflict and determine methods by means of which these conflicts may be overcome.

Conclusions

In this chapter we have tried to indicate certain trends of modern psychology. Realizing the necessity of dealing with the individual as a whole, experimental psychology has shifted its direction toward individual psychology. To understand the individual as a whole it has become necessary to study him in his social environment. The social psychology of the past twenty-five years has been mixed up in many tangling alliances with sociology and other social sciences, but it is gradually finding itself. Its data and problems are beginning to emerge in a more clean-cut fashion. It is the conviction of the writer that the division of the field into "differential psychology," on the one hand, and "integral psychology," on the other, is a valuable division for the progress of social psychology as a whole. It is only recently that social psychology has recognized that the study of individual differences constitutes its most valuable data. These differences have been described qualitatively, and many of them have been measured quantitatively. These quantitative studies are tending to make social psychology more and more an exact social science.

This chapter has been written on the assumption that the place of the adult in the community is determined and defined by the way in which he differs from others. His deviation from the mean or average of his group in such variables as bodily size, shape, physiological functions, and psychological processes and abilities, as well as in character, temperament, and personality, are the all-important factors. While many human variables have been isolated and measured and intercorrelated in many ways, yet the work is scarcely begun. The experimental social psychologists of the future will spend much of their time and energy in studying and measuring individual differences.

But differential psychology represents only one side of social psychology, just as differential calculus is only one part of that discipline. The other side (to follow the suggestion from the calculus) is "integral psychology." Modern social psychology assumes that individuals are "integrals" and not "integers" (30). The problem of "integral psychology" is that of integration or finding out how an organism or a community functions as a whole. This process is traced genetically from the lower organisms to complex human communities. Three major principles of integration are found: communication, control, and drive or motor force. For any complex machine or assemblage of parts to work together as a whole, energy must be communicated to the parts in some systematic and controlled way. A society must have a drive and a direction; it must have communication between parts; and there must be means of control.

REFERENCES

1. Allport, F. H. Social psychology. Boston: Houghton Mifflin, 1924.
2. Alverdes, F. A. Social life in the animal world. New York: Harcourt, Brace, 1927.
3. Bingham, W. V., & Freyd, M. Procedures in employment psychology: a manual for developing scientific methods of vocational selection. Chicago: Shaw, 1926.
4. Burt, C. The young delinquent. New York: Macmillan, 1925.
5. Cattell, J. McK. [Ed.] American men of science. (3rd ed.) New York: Science Press, 1921.
6. Chapman, J. C., & Wiggins, D. M. Relation of family size to intelligence. *Ped. Sem.*, 1925, **32**, 414-421.
7. Child, C. M. Physiological foundations of behavior. New York: Holt, 1924.
8. Child, C. M., *et al.* The unconscious: a symposium. New York: Knopf, 1927.
9. Clarke, E. L. American men of letters. New York: Longmans, Green, 1916.
10. Cleeton, G. E., & Knight, F. B. Validity of character judgments based on external criteria. *J. Appl. Psychol.*, 1924, **8**, 215-231.
11. Cooley, C. H. Human nature and the social order. New York: Scribner's, 1902.
12. Cox, C. M. Genetic studies in genius. Vol. 2. Stanford Univ.: Stanford Univ. Press, 1926.
13. Dashiell, J. F. Fundamentals of objective psychology. Boston: Houghton Mifflin, 1928.
14. Ellis, R. S. The psychology of individual differences. New York: Appleton, 1928.
15. Freeman, F. R., *et al.* The influence of environment on the intelligence, school achievement, and conduct of foster children. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928, Pt. 1, 103-218.
16. Fryer, D. Occupational intelligence standards. *School & Soc.*, 1922, **16**, 273-277.
17. Galton, F. Hereditary genius. (2nd ed.) London, 1892.
18. ————. Inquiries into human faculty and its development. London, 1883.
19. Gowin, E. B. The executive and his control of men. New York: Macmillan, 1925.
20. Hartshorne, H., & May, M. A. Studies in deceit. New York: Macmillan, 1928.
21. Healy, W., & Bronner, A. F. Delinquents and criminals. New York: Macmillan, 1926.
22. Hollingworth, H. L. The psychology of thought. New York: Appleton, 1926.

23. Judd, C. H. The psychology of social institutions. New York: Macmillan, 1927.

24. Kantor, J. R. What are the data and problems of social psychology? *J. Phil.*, 1923, **20**, 421-456.

25. Kitson, H. D. The psychology of vocational adjustment. Philadelphia: Lippincott, 1925.

26. Kretschmer, E. Physique and character. New York: Harcourt, Brace, 1925.

27. McDougall, W. Social psychology. (8th ed.) Boston: Luce, 1914.

28. Murchison, C. Criminal intelligence. Worcester, Mass.: Clark Univ. Press, 1927.

29. ————. Social psychology: the psychology of political domination. Worcester, Mass.: Clark Univ. Press, 1929.

30. Ogden, R. M. Psychology and education. New York: Harcourt, Brace, 1926.

31. Paterson, D. G. & Ludgate, K. E. Blonde and brunette traits: a quantitative study. *J. Person. Res.*, 1922, **1**, 122-127.

32. Pintner, R. Intelligence testing. New York: Holt, 1923.

33. Slawson, J. The delinquent boy: a socio-psychological study. Boston: Badger, 1926.

34. Terman, L. M., *et al.* Genetic studies in genius. Vol. 1. Stanford Univ.: Stanford Univ. Press, 1925.

35. Thomas, W. I., & Thomas, D. S. The child in America. New York: Knopf, 1928.

36. Thorndike, E. L. Intelligence and its uses. New York: Harpers, 1920.

37. Warren, H. C. Human psychology. Boston: Houghton Mifflin, 1919.

38. Watson, J. B. Psychology from the standpoint of a behaviorist. Philadelphia: Lippincott, 1919.

39. Wheeler, W. M. Social life among the insects. New York: Harcourt, Brace, 1923.

40. Wiggam, A. E. Exploring your mind. Indianapolis: Bobbs-Merrill, 1928.

41. Woodworth, R. S. Dynamic psychology. New York: Columbia Univ. Press, 1918.

42. ————. Psychology. New York: Holt, 1921.

43. Young, K. Source book for social psychology. New York: Knopf, 1928.

CHAPTER 20

THE CONFLICT AND SURVIVAL OF CULTURES

Clark Wissler
Yale University

A subject like the conflict of cultures, approached from a psychological point of view, calls for an integration of several sciences. First, social psychology has something to contribute both as to method and data. Secondly, for purposes of control, social anthropology must be appealed to. Finally, the traditional social sciences, history, sociology, and economics, must be taken into account. We cannot hope to make full use of the aids and materials these different sciences offer; no living individual is master of them all; but we can draw upon them for data and points of view. If, for example, we wish to know how other peoples than our own carry on their social affairs, we must draw upon the descriptive literature of anthropology and history, as the case may require. It is usual to turn to anthropology for examples of fundamental social behavior, the assumption being that among primitive tribes of men will be observed the elementary forms of social life. That the lives of such tribes are less complex than ours is obvious, but it is not well to take for granted that in experience and attitude the peoples dealt with in anthropology are representative of the original state of man. While it is true that anthropology aspires to be the natural history of man, anthropologists find living tribes following a mode of life far removed from what is conceived to be the primitive form of society. We can, however, safely begin with observations upon peoples as they are found and note how they function under existing conditions, savage and civilized.

There is in the making a sizable body of opinion concerning the psychology of primitive peoples. It is not necessarily to be assumed that the processes to be observed in primitive life are different from those operating in civilization. Yet something is gained by focussing attention upon primitive cultures, since we are not a part of them and for that reason can the more easily deal with them in an objective way. On the other hand, a comprehensive psychology of society should embrace the civilized as well as the uncivilized. Most of what is now called social psychology is based upon observation of the individual responding to social situations. However, it is well to bear in mind that such a study is far more complicated than the simple statement might indicate. To follow through an investigation of this kind, one must regard human nature, the social aspects of personality, the motives that express themselves in social relations, and finally attention must be given to culture itself. While students of the social sciences are interested in such studies, they usually regard social psychology as a division of psychology, rather than of the social sciences. This is significant to our inquiry because if accepted as our point of view, it implies that problems in culture conflicts are to be approached through the individual by accepted psychological technique.

Yet before deciding upon an expedient method of approach, note should be taken of the various points of view assumed by sociology. The history of sociology suggests that the first important step in the modernization of this study was the use of biological technique, or the adaptation of biological concepts. For example, a society was conceived as an organism, not homogeneous throughout, but an integration of structures and functions sufficient to secure unity of action. Later on, this idea was displaced by the notion that society was essentially a psychological phenomenon and that psychology was the key to the problem. As experience in sociological research accumulated, however, it developed that there was a sociological approach to the problem, distinct from the psychological, and that, in so far as there was a field for social psychology, it was with psychology rather than with sociology or anthropology. On the other hand, it should be recognized that social psychology is still in the pioneer stage; psychologists are feeling their way into primitive cultures and modern social problems. The present subject-matter of social psychology is comprehended under descriptions of social consciousness, observations on the responses of the individual to social situations, both natural and experimental. Also, the term "social behavior" is used to indicate the psychological approach to culture.

The consensus of opinion seems to be that, if social psychology is to become anything more than a descriptive science, it must proceed by observation and experiment. As we noted at the outset of this discussion, most treatises on social psychology draw upon data obtained by anthropologists, in the hope that by comparing the responses of individuals in primitive social situations with those confronting ourselves, fundamental processes and tendencies will appear in clear relief. In the main, however, the results of such comparisons have been disappointing. This might have been anticipated, since the data offered by the literature of anthropology are rarely strictly comparable to that for contemporary civilization. The latter is subject to many checks for accuracy; but what is still more significant, it can be evaluated because of our intimate relation to the order of life in which it finds its setting. We are not a part of uncivilized society and, therefore, prone to misinterpret and unduly weight the isolated observations recorded by travelers and anthropologists. Since the social psychologist is essentially a psychologist, he approaches the data chosen from the standpoint of the responding individual; whereas the anthropologist pays little attention to the individual, fixing upon the group or the community as a social whole. It is true that the anthropologist may record events in the group, involving a part of the whole, as only the men, the old men, a band, etc., but it is only such of these events as relate to the activities of the group as a whole that he regards as of importance. For example, we may be told how rigid the force of a taboo is held to be, recounting how everyone in the group concerned is expected to be constantly on the watch lest the taboo be violated. What we are not told is, as to the attitude of the individual, how faithfully he observes it, whether he is always in a state of fear, or whether he regards the taboo as we do an automobile when crossing a street. Had the social psychologist made the observations, we should expect information as to how the individual responded. Nevertheless, anthropology does furnish data which may be profitably viewed from the psychological angle.

However, before venturing into the domain of primitive man, we must take into account the conceptions and methods of the anthropologist. First, he notes that viewed in the large, man is a community dweller. The camp and the village are universal throughout the civilized world and even in modern times urbanization is the outstanding characteristic. The ultimate social unit for the anthropologist is not the family but the community; it may so happen that all the members of a community are related by blood, but that is incidental to the phenomenon of community life. From the point of view of the social psychologist, these communities function because of the cooperative responses of the individuals composing them; as these psychologists conceive, it is just because humans are disposed to cooperate readily and effectively that a community is possible. The anthropologist looking in from the outside notes that such a community follows more or less standardized procedures, which in the older literature were listed as manners and customs, but to which the technical name culture is now applied. Anthropologists prefer to work with detailed accounts of community cultures and these facts make up the data of what is sometimes called social anthropology. The term "culture" as used in social science means an aggregate of customs which is theoretically resolvable into culture elements.

This may be clearer if we take note of how the anthropologist works. He begins his study of tribal culture by concentrating upon one or two points. Thus, he may observe how fire is kindled, noting that it is made by boring one stick into another, but that these simple-looking implements are fashioned according to a specific tribal pattern and that the procedure is likewise fixed. Yet, it is not enough to say that fire is kindled by wood friction; the individuality of the implements and the accompanying procedures must be recorded and representative objects collected. Thus, the fire-making implement, accompanied by photographs and field notes, becomes the objective record of a unit of observation. Such is, for practical purposes at least, a unit of the tribal culture and is spoken of as a *trait*. This term is also applied to mannerisms and to customs of whatever kind. Thus the custom of a man marrying his wife's sisters may be observed and, if so, is set down as a trait of the tribal culture. It follows then that a tribal culture, as conceived by an anthropologist, is characterized by the enumeration of its observable traits and that the culture of one tribe is distinguished from that of another by differences in these traits.

While theoretically the ultimate element in a culture is a trait, it is seldom possible to carry one's analysis so far. Even in case of making fire with a wooden drill, there is some difficulty in deciding whether the object, the drilling to make fire, the memorized technique, or each of several other factors constitutes an element. Because of this difficulty it is customary to use the term "trait complex" and speak of the simpler complexes as traits and the more complicated as trait complexes. Thus the "ballot" in our culture is a trait complex because it involves many elemental procedures. For a fuller understanding of the anthropological concept of culture the reader should turn to the literature of that science. However, it will suffice here to note that in dealing with culture traits the social anthropologist begins with what the community has produced, the end result rather than the processes concerned. The writer has in mind a certain community in which all hasten to perform or participate in a fixed ceremony at the first sound of thunder in the spring. This procedure may be called a trait complex; if we inquire as to the origin of such a custom, we may be told that it is the outcome of social and other processes, and is the end result of a development. The social processes involved are vested in individuals and are largely psychological, but the anthropologist rarely goes back of the trait complex, or the concerted action of the group. He accepts the formulation of the custom as carried in the memories of community members; in the instance just cited this formulation might be, "at the sound of thunder in the spring everyone is to respond thus." This makes clear the distinction between the interests of psychologists and anthropologists; the psychological problem lies in the response of the community rather than in the formalized precept or technique.

On the other hand, the anthropologist does take note of the attitudes of one community toward another, what procedures they have formulated for contact with, for protection against each other, etc. Because he notes that communities living as neighbors have many traits of culture in common, he suspects that they have influenced each other, in much the same way that individuals react to one another. In the clinical aspects of psychology the term conflict is used to indicate lack of adjustment between the individual and his surroundings, and in much the same way communities are said to experience culture conflicts when they are not well adjusted to each other. Such conflicts arose when the American Indians were confined to reservations and pressure brought to bear upon them to adopt white standards of living. An immigrant community of Poles, Finns, etc., taking up residence in the United States, Brazil, or elsewhere, may observe conflicts between the cultures they brought from the home land and that of the country in which they now live. It seems best in such cases to distinguish between the conflict or antagonism of culture traits or patterns and the responses of the individuals concerned. The psychological interest is primarily with the latter, but, since the response is usually provoked by a situation, it seems necessary to study the situation as preliminary to the problem presented in the responses of group members.

Social Life a Community Affair

It may be advantageous to consider the conditions necessary to social phenomena. The existence of a culture is dependent upon the normal functioning of a community. The part of the individual in a community is considered elsewhere in this volume, our present interest being directed to the community unit in its relation to other similar units. One pertinent question is whether a single community can be sufficiently self-stimulating to develop a culture in the anthropological sense of that term. This need not be taken as an analogy to the relation of one individual to another, however true it may appear that our mental life is dependent upon the interaction of individuals. The objective data of anthropology is in itself sufficient to put the question. One of the accepted axioms in anthropology is that the culture of a group is the result of give and take among neighboring groups. At present, however, we are considering the group or communitiy as a culture unit. While theoretically two persons may be sufficient to initiate social responses, they would not constitute a community in the anthropological sense. It is assumed that for each given time and place there is an optimum size for a community to function normally. In actual count of individuals the community population may show a high variability around a mean. Unfortunately there are no reliable statistics on primitive populations, but villages and camps as reported among such peoples are rarely large. In any case it is reasonable to expect that the richness of community culture will bear a relation to the number of individuals involved. Oversize and undersize are certain to interfere with efficiency in community functioning.

Communities are often federated to form a large political unit to which the name tribe is given. Tribes, as so defined, usually maintain a uniform language and culture. For example, such familiar Indian names as Cheyenne, Omaha, Navaho, etc., designate tribes in which there is uniformity of speech and culture. Each of these tribes, however, is composed of communities, the primary culture groups. These distinctions are to be kept in mind because when we turn to the problems of culture conflict it is usually tribe against tribe, whereas when we take up the interaction of groups, the phenomenon appears at its best as between the communities within the tribe.

Community life is no more static than is the life of the individual. Change is the natural order of social life. When communities are observed, they appear to be developing differently or manifest what has been called group-difference tendencies. A new individual joining a community finds it following well-formulated customs to which the community members respond more or less in unison. As an example of these customs, note the aforementioned procedures at the sound of thunder. How the group arrived at this procedure we do not know, but the anthropologist answers that it is a resultant of individual responses and circumstances peculiar to the place and time. Since this custom seems typical of most culture elements, the anthropologist assumes that all culture elements originate in this way. Further, it is assumed that the responses of the community member will vary from generation to generation and thus bring it to pass that many elements of culture change. So it is assumed that all groups will present different tendencies; the communities constituting a tribe will differ somewhat, the tribe in turn differ from other tribes, etc. What the anthropologist observes, however, is that the differences within the tribe are but slight, while those between tribes present a wide range of difference.

Social Behavior

Considered zoölogically, man is a species and as such possesses a form of behavior peculiar to the species; we speak of this as human behavior. Social phenomena are considered manifestations of human behavior. All living peoples, civilized or savage, seem to possess the essentials of this behavior, at least each known community operates with a standardized speech, uniform customs, procedures, etc.

This assumption is taken as the point of departure in a psychological approach to primitive culture, that is, we assume that primitive man is capable of the same kinds of responses as are observed in civilized man. The problem may then be stated as observing what responses primitive man makes to specific situations. Thus in psychology the procedure is to classify mental responses and focus attention now upon one type of response and now upon another. For example, sensation, perception, memory, thought, etc., can be investigated as if they were so many separate processes in the life of an individual. Yet these processes are assumed to integrate in the individual in responding to social situations. Further, in the life of uncivilized man these same processes are assumed to operate. We begin our study, therefore, with the expectation that merely to analyze a social act and define it in the terms of traditional psychology, or to translate it into terms of individual response, will give about the same answer in respect to primitive society as civilized. In other words, we assume that the individual everywhere in similar social situations is capable of the same responses. Anthropology on the other hand demonstrates that the various groups of primitive peoples maintain different institutions, conventions, and techniques, which may be taken as the results of social activity. Yet, Bartlett (2) has demonstrated the futility of attempting to explain the origins of these traits of culture in terms of individual response. Social behavior implies a group with its accepted behavior patterns; anthropology makes clear that these patterns not only differ as between primitive and civilized groups but that primitive groups differ among themselves in this respect. Accepting such group culture patterns as objective phenomena on the one hand, and recognizing the individuals of the group on the other, the psychologist is ready to face his problem. One immediate aspect of this problem lies in the responses of individuals to the culture patterns confronting them.

Culture Determinants

Now that we have stated the points of view taken in this study, we may give further consideration to origins and determiners in culture. In the older literature of anthropology, especially the works of Tylor (13) and Frazer (4), origins and determiners for culture traits were sought in mental response. For example, Tylor accounts for the belief in souls in terms of individual experience. Finding that all known peoples, however primitive, believe in souls, in a material and a less material existence, he put the question as to what individual experiences could lead to such beliefs. Though he does not say so specifically, Tylor seems to have assumed that all known savages could respond in the same manner as a European, i. e., he took it for granted that the mental constitution of mankind was the same everywhere. His method was to explore human experience for situations suggestive of such a dual existence as a belief in souls seems to imply, finding in sleep, dreams, imagery, etc., not only provocation that seemed adequate but situations in which, human nature being what it is, no other conclusions seemed possible. According to Tylor, the inert body of the sleeper, his obliviousness to what was going on around him, the recollections of dreams by the sleeper indicating that some part of himself went on a journey, leaving the body behind, all suggested a dual existence. In other words, Tylor conceived that all primitive people, daily subjected to these experiences, would respond in the same general way. So, according to the terminology of anthropology, the culture trait complexes of beliefs in souls, found everywhere in a multiplicity of details, are based upon a universal response. Stated in behavioristic terminology, this would mean that belief in souls is determined by innate behavior. Tylor recognized the part played by tradition in spreading and perpetuating such beliefs, once they were formulated, but insisted that sooner or later every community would formulate them independently.

This is not the place to review the controversies that have arisen in response to this challenge of Tylor. Some critics regard the hypothesis as too rationalistic; others consider that such a belief was not spontaneous, but arose somewhere in the accidental responses of an individual, was adopted by his tribe and then passed on to other tribes, ultimately spreading over the whole inhabited world; the socially minded object to the theory because they regard all culture complexes, of which belief in souls is a typical case, as naturally emerging from group situations rather than from individual responses. Our interest at present, however, is not to attempt what no one has as yet achieved, the solution of the problem, but to call attention to its psychological aspects.

The problem of belief in souls which we have just cited is typical. The difficulty lies in that no one can say where and when primitive men did first respond in the manner assumed. We can, however, face about and, accepting the fact that the belief in souls is universal, consider this culture complex as something to which individuals respond. We shall thus have met one of the objections to social psychology, viz., the futility of search for causal sequences between individual responses and culture patterns. In such a case the individual is not lost sight of, but is placed in a contemporaneous relation to the patterns to which he responds. It is conceived that, though causal relations between the initial individual response and the resulting culture pattern may not be recoverable, yet data upon the sequence from pattern to response are readily observable and subject to check.

For example, the culture complex of incest avoidance is an intriguing problem in sociology. In a well-ordered community there is a standardized pattern for behavior in this respect. There is, for one, a classification of individuals in terms of relationship, thus defining the incest group with respect to each individual. Anthropologists observe that the classification varies among primitive groups but that the distinction is always drawn. The recognition of incest is, therefore, universal. We can now study the response of the individual to that recognition and its relation to other responses. On the other hand, we may seek the origin of the culture pattern in individual response. It is still an open question as to whether incest is a spontaneous response in the individual or an induced social phenomenon. It may be both, a basic natural antipathy to house-mates on the one hand, which may be accentuated or inhibited by a culture pattern. The solution of the crucial problem may be advanced by studying the behavior of the individual subjected to a specific pattern of incest avoidance, supplemented by studies in animal behavior.

Respecting the determiners of culture trait complexes, then, it appears that the problem is open to the psychological approach. Culture patterns are conceived as determiners of individual responses on the one hand and as the results of community response on the other. The probabilities, however, favor the origins of culture traits in spontaneous community response, and it may be helpful to speculate upon them for a moment. We have mentioned the assumption of Tylor that there was an animistic response. Genetic psychology has also made a good case for this in the behavior of children, though this still needs to be checked by experiment. Turning again to culture, one of the most universal responses is the recognition of sex distinctions. Convention has failed to suppress the automatic recognition of the two classes, presumably because bodily characters and behavior are ever present; the response is a recognition of two classes in the community. With two classes so inevitably fixed and perpetuated, recognition of which is spontaneous, the way is open to all the conventional class divisions of function and responsibility observed in the community. Anthropologists have emphasized the want of consistency in the division of labor between men and women in different cultures. Thus, in some instances, women spin and make the clothing; in others, the men perform this service, such differences in custom being taken as proof that all sex distinctions are idiosyncrasies of behavior. It is necessary to distinguish however, between the substance and the form. What is universal is the sharp line drawn between the two sex classes with respect to many activities. Every people known recognize a number of privileges and obligations peculiar to each sex alone, but the specific privileges and obligations vary. It is then the recognition of sex classes that constitutes the behavioristic basis to this phase of culture and this in turn seems spontaneous and inevitable. Further, since the basis of group life is the differentiation of functions not so much of individuals as classes, it is not strange that functions other than those relating to sex should be placed in secondary association with a sex class. Passing on to other examples of response, we note that almost equally spontaneous is the recognition of the mother as the feeder of the child, and curiously enough, all the world over, the woman has borne the brunt of preparing food for the man. In this case it may well be that the recognition of the primary function has controlled the secondary one. The automatic recognition of the mother as the feeder puts all women in a class with whom that function is associated, in contrast to men.

Much speculation has been given to the widespread practice of menstrual ceremonies. First of all, we should take note of the objective aspect of the phenomenon; when the event occurs, the response of the community automatically puts the girl into a class, or moves her from one class to another. She is now an adult woman. The keeping of age records in years is a modern procedure, quite foreign to primitive cultures; hence, menstruation is one definite age marker, thus enhancing its weight as a stimulus. The response of the community is formalized in a variety of procedures.

Linguists frequently call attention to the unity of speech in a primitive group. Each such group possesses more or less individuality in speech, resulting in such classifications as dialects, languages, and stocks, according to the degree of difference. It appears that differences in speech are quickly noted; the response is lively. A stranger approaching is quickly classed as a foreigner if his speech is different. He may even speak the language, but his strange accent is sufficient to put him in his class. Such responses are universal and will occur over and over, no matter what the social ideal may be. We are not now concerned with what kind of culture patterns communities set up, but what they respond to. Speech, sex, sleep, death, food, etc., are universal and every people will respond to them; in some cases these responses are uniform. These uniform responses constitute in part, at least, the psychological basis for customs. They also serve to conserve the form and scope of community customs and are largely responsible for culture following one general pattern.

Reaction of Cultures

Anthropology shows that in the primitive world the tribes living in a geographical area are usually well informed as to the culture characteristics of each other. They feel a certain emulation just as do communities in modern life. An American town will apologize for not having a library, an art museum, etc., because these are prominent assets in neighboring towns. There is often a definite feeling of inferiority in not possessing them. With this the reader is familiar. A citizen of one town may know nothing of art and care less, may never have read a book in his life, and yet lead a movement by which both an art museum and a library are added to the equipment of the town, nor is there anything abnormal in such a procedure.

Primitive communities react in similar ways. One tribe will apologize to a traveler for not having a great ceremony like their neighbor's. They would acquire it if circumstances were favorable. They look up to their neighbor as a leader, a model. If strong and virile, the tribe will emulate: if weak, feel its inferiority. One of several courses may be pursued, the leadership may be accepted and the culture of the dominant tribe participated in as far as possible; on the other hand, resistance or rather resentment may be offered, leading to strife, etc. The culture area of anthropologists is the habitat of a group of tribes having more or less similar cultures. The similarity of these tribes is accounted for in anthropology as due to borrowing. What is observed, however, is that two or three tribes dominate, their cultures are richer, and they stand out as the center of influence in the area. These leading tribes realize a high degree of self-assertion; the less typical tribes live more in the shadow, they are free and independent, it is true, but they are aware of their powerful, prosperous, expanding neighbors, who lead in culture. In brief, a culture type appears as a mode of life evolving in centers or among a cluster of tribes and more or less distributed among the outlying neighboring tribes.

BORROWING

One of the processes upon which anthropologists now put great store is that of borrowing customs and practices, or the contact of cultures. If a tribe devises a different way of kindling fire, a new decorative design, composes a new song, etc., neighboring tribes tend to adopt them. This process has been explained in psychological terms as imitation. One tribe is said to imitate another, and so it comes about that neighboring tribes have closely similar cultures. Yet, if we analyze the procedure, we see first of all a spontaneous response, a lively interest in the ways of the neighbor, which may lead to the acquisition of these individualities. This may proceed quite naturally without much in the way of reflection or rationalization. On the other hand, borrowing may be rationalized and deliberate, as when a people attempt to enrich their culture by seeking and choosing from the cultures of others. However, an objective comparison of tribal cultures suggests several types of borrowing, as:

1) taking over isolated elements of culture.
2) acquiring culture complexes.
3) taking over entire sections of culture, as economic system, religion, etc.

Mythology furnishes illustrations of the first, as when a new story is learned from a visitor to the community. The casualness of such borrowing is noticeable. Most tales deal with the responses of individuals to social situations and so are directly intelligible to the hearer; to repeat it calls for no new equipment.

A magical belief may also be borrowed in much the same way. If one hears that "marking on the ground with a stick will cause death," he may circulate the statement in the community, where it may be accepted as a probability and so evolve into a standardized belief. The Indian substituted the colonist's steel knife for his own stone knife, the gun for the bow, etc., all of which are examples of culture borrowing, trait by trait. Many other illustrations could be added, in all of which the members of the community spontaneously take over the trait of a neighboring tribe, or even acquire it from a chance visitor.

Trait complexes present a more involved situation. If such a complex is borrowed, many related procedures must be learned and all of them adjusted to the old order of life. A good illustration is the domestication of the horse by the American Indian of our western plains. The Spanish settlers in Mexico introduced the horse to the Indian tribes with which they came in contact, taught some of them how to handle and care for these animals. A true culture complex was involved in training, care, feeding, breeding, riding, etc. To make the horse an important acquisition the whole of the technique must be learned or invented anew. A study of our Indian horse culture indicates that nothing of importance was invented by the Indians; they merely followed the pattern set them by the European settlers. Once, however, the tribes nearest the settlements learned how to handle horses, the neighboring tribes out in the open country acquired the complex in turn, and so it was relayed along even in advance of exploration. It is true, however, that Indians had used dogs for transporting baggage and so possessed patterns of domestication and animal transport before they came in contact with this new complex. This may have offered a situation more receptive to acquiring the horse complex, but the obvious advantage of horse transportation to the Indian would doubtless have urged him to borrow the complex in any event. Nevertheless it should not be overlooked that in this case conditions were favorable to the borrowing.

While the horse complex was borrowed quickly, some other complexes in our culture have been taken over by the Indians slowly and imperfectly. Thus one of the more important trait complexes in European culture is that of individual ownership of land. This is a true complex because it is not merely the idea of seizing a piece of land that is involved, but the system of regulations by which ownership is maintained, protected, and transmitted. Few primitive tribes possessed such a culture complex and found it more difficult to borrow than the domestication of the horse. In the case of the horse, little more was involved than the adding of a new facility for transportation, not unlike substituting the automobile for the wagon. In the matter of land ownership it was otherwise. The Indian of our western plains did not entertain the idea that the land could be owned by an individual; to him it was the range of the tribe in which all had a share. Individual property in land, with all its complications, was a new idea. Further, he was a hunter and not an agriculturist; hence he felt no need for this new trait. So he did not take to land holding spontaneously; it has been necessary to force this trait upon him.

The third form of borrowing is that in which a series of complexes is taken over, or a culture as a whole. ·Changing the economic base to a culture would be an instance of such culture change, or the shift from a hunting to an agricultural mode of life. The adoption of the Christian religion and its closely interrelated ethics and customs is a more familiar instance. History records how a native Hawaiian king sent to America for teachers of the new religion and by edict overthrew the old native religious practices, substituting the new. While it is true that time was needed to make the change, it was made without unduly disturbing other phases of life. We could complete the picture by citing examples of the borrowing of an entire culture. Students of primitive tribes often call our attention to the existence of tribes speaking related languages but practicing different types of culture. In such instances it is clear that a complete change in culture has taken place. One of the classical examples is that of the Hopi Indians who practice agriculture, live in well-built towns, have a religion like that of other southwestern Indians, but at the same time show by their language relationship that they were once a wild hunting tribe.

The Procedure in Borrowing

It is hoped that the foregoing illustrations may serve as the background for a discussion of the procedure in borrowing. First, let us consider the acquisition of a single trait. A single individual may introduce the new trait. It is also obvious that such a trait may be his own invention and not something he observed in the culture of a neighboring tribe. The bow, for example, is a mechanical device, which we may safely assume to have been invented by an individual, probably first as a toy. The first step having been taken, or the invention made, its socialization would be the next procedure, that is, the spread of its knowledge, manufacture, and use throughout the community. Thus, from a functional point of view, there is no essential difference between the innovation of a single trait and the initial borrowing; the process seems to be the gradual spread of the idea or the practice among the members of the community. On the other hand, the chances of success seem greater in borrowing, because the inventor may fail to introduce his new device and thus his contribution to culture be lost, on the other hand, if he borrowed the idea from the culture of a neighboring tribe, others will observe it also and thus greatly increase the possibility of its introduction. There is still another factor to be considered. Few traits are the work of an individual; the initial step is individual, but, increment by increment and refinement by refinement, the trait takes definite form. This is a long process and there are many chances for the growing trait to fail; on the other hand, once a tribe has perfected a trait, it can be taken over without loss of time. The economy of such a procedure is obvious. Finally, the procedure resolves into individual borrowing, learning to use the new trait, or to follow the new custom.

The borrowing of complexes appears more distinctly the work of a leader and his followers. This is particularly true in the case of cults and ceremonies. For example, a Blackfoot Indian visited a neighboring tribe where he learned a ceremony; upon his return he constructed a ritual similar to the one he had observed, added a few original but minor features, and then began to collect followers. As in all such cases, his ritual and his teaching were supposed to enhance the well-being of the devotee. In a few years there was a respectable minority in the tribe believing in and studying the tenets of this cult and so it came about that this complex of beliefs was recognized as a part of the tribal culture. Such borrowing of a complex, in contrast to the taking-over of single traits, cannot be considered spontaneous. Rather is the conscious effort of an enthusiast necessary to secure and maintain a following. This does not imply that there is active resistance to the new complex on the part of the group; such resistance there may be, but, since a complex of new ideas and procedures is involved, time and energy are necessary for their acquisition. Here also are opportunities for class distinctions, as the clever, the rich, the old, the young, etc. If the new complex is intricate it may be accessible to the clever only; if property resources are requisite, then the rich are favored, etc. Yet eventually, the complex will filter down or up as the case may be, until it comes to be a common culture heritage.

Culture Change

Another aspect of our subject is culture change (8). Even the culture of a primitive tribe, as in Australia or New Guinea, is not wholly stable; changes seem to be under way always. These changes may result from both internal and external stimuli; in one case the new trait or complex is borrowed from a neighbor, in the other it is invented from within. But change in culture is not wholly dependent upon these; often there seems to be an unconscious drift or development. Such drifts have been noted in linguistics, as the progressive changes in phonetics and structure. Also, it seems possible that interest in a culture complex cannot be long sustained if there are no changes in it; that is to say, the response may be spontaneous to certain variations in the procedure, and such accumulative selections cause a series of changes the ultimate outcome of which might well be a different trait complex. It is important to bear this in mind in this discussion, since conflicts of culture are not necessarily collisions between two static systems but may occur between drifts or directions of change.

Group Dominance and Leadership

The rivalry of culture groups can be studied by the use of anthropological data. The culture areas to which we referred at the outset are best explained as the fringes to centers of culture influence. The situation here is, however, not essentially different from that of a modern nation leading economically, intellectually, or artistically, as the case may be; all the surrounding nations do their best to follow. In this case the effort may be largely socially directed and in so far a conscious process. On the other hand, the efforts of primitive tribes to acquire the culture of the dominant tribe is often conscious. We have the example of the Nez Perce Indians sending a committee to the frontier settlements along the Mississippi to invite a white missionary and teacher to give them the elements of the superior culture. Other instances of such initiative are on record, and the traditions of primitive folk often recount how one or more of their number went abroad to learn new traits of culture.

Looking at groups in general, we see that they not only recognize differences between themselves and other groups but are conscious of their own position. If they are the dominant group, they are aware of that; if not, they recognize the superiority of their neighbor and attempt to emulate him. In other words, they are conscious of their relative position whether they admit it or not.

Conflict of Cultures

The foregoing may serve as orientation for the consideration of what may happen when two contrasting cultures come into contact. In such situations, single traits may compete with each other. Examples of this are seen in the displacement of the bow by the gun, the fire-drill by the match, the earthen pot by the iron kettle, etc. There is a minimum of resistance in such cases, but indirectly the displacement may jar the whole tribal system of procedure. It is easy to throw down the bow and take up the gun and to learn to kindle fires with matches; but this new equipment cannot be produced within the tribe, as heretofore. The tribe is now dependent upon the trader. Further, in order to secure these objects, other goods must be produced in sufficient quantity, perhaps furs. This seems simple enough, but standards of individual property are involved, social obligations must be adjusted to the trapping calendar, etc. Success in trapping becomes more important, so the magical beliefs of the tribe must be adjusted. Nevertheless, such conflicts as arise are rarely serious unless some strong taboo is violated. Thus it is said that some Hindu troops once mutinied because the paper cartridges, whose ends they were required to bite off, were greased with an animal fat, against which there was a taboo. But the obvious ease of adjustment in this case may be taken as typical and indicating that, on the whole, no serious conflict need be anticipated from the introduction of single traits.

Before leaving this subject, it is well to be reminded that not all introduced traits are dependent upon foreign trade. Thus the horse was introduced to the American Indian, who propagated and equipped the horse by their own labor; the tribe was thus in no wise dependent upon a foreign source of supply. Economically, the tribe was enriched, but no important adjustments to other culture traits were necessary. That there was a psychological reaction is probable, since this new mode of transportation gave a feeling of power, increased geographical knowledge, and gave more leisure. The tribe receiving horses might be expected to feel superior to neighboring tribes and the reverse. Finally, new information may be acquired as to distant lands, as to the properties of natural resources, etc., thus increasing the store of knowledge. A new tale or myth may be acquired, a new song learned, a new game adopted, all without disturbing existing traits. So in general the introduction of a useful domestic animal, a plant, a new industrial art, or a new fact will enrich a culture with a minimum of conflict.

On the other hand, trait complexes displacing other complexes seem to present serious situations. Yet these seem to vary according to the relative values attached to the complexes involved. An example of conflicts of this kind are those which arise respecting individual property and thrift. Most primitive peoples feel that every individual is entitled to food and other necessities, whether he is a producer or not. They do possess the idea of individual property, but entertain an ideal of hospitality which requires frequent sacrifice to others. Lewis H. Morgan (7) first sensed the significance of this aspect of primitive cultures, designating it as the law of hospitality. When nations are subject to European control, this tendency conflicts with the system of individual economic independence. For example, the younger members of the tribe may be educated and induced to undertake farming, but when the crop is harvested, the old law of hospitality requires that the producer give it away. To refuse to do so would be grossly immoral from the point of view of his tribe. Such a conflict of culture complexes may be very distressing but of a less emotional character than some others.

Among all primitive peoples the distinctions between the work of men and women are sharply drawn. To disregard these distinctions is to stigmatize oneself. Among some tribes all preparation of food is not only the work of women but they also own the food; under such circumstances, for a man to claim ownership in food would be disgraceful. In most languages, the strongest terms of contempt are those describing the individual in characterization of the opposite sex. Yet while there are everywhere distinctions between the work of men and women in the tribe, these distinctions may vary as to the kind of work. Among certain hunting tribes of Indians, it was woman's place to tend the fields, man's place to hunt. The customs of white culture, more dependent upon agriculture, reversed the distinction. So when Indian and white culture came in contact, the Indian resisted assuming responsibility for work in the fields as shameful and humiliating. When hunting was no more possible, the white man insisted that the Indian turn to agriculture; the conflict resulting and its emotional intensity are obvious. Humiliation and degradation were the price the Indian must pay.

Still other examples may be cited. Among many tribes, the place in which a person dies is taboo; thus, if a person dies in a house, the place will be deserted, to fall into ruins. With the simple shelter of primitive peoples, this custom was easily maintained, but when the European custom of more substantial and expensive buildings is brought in, the difficulties increase. Fear of spirits, bad luck, the whole background of beliefs concerning the dead, all combined to make life in such a house unbearable. Their feelings would be about the same as those of a European asked to live in a cemetery. But on some of our Indian reservations there is still further conflict. To provide the Indians with adequate medical care, hospitals have been built; the difficulties in their use by the Indians are obvious. Once a patient dies in the hospital, it is no place for a sick person. The violation of the taboo, the emotional effect upon the patient, if forcibly taken to the hospital, all tend to unite the tribe in opposition. Suspicion may follow as to the motives of the white race in inducing them to use the hospital, and this crystallize into a belief that the whites desire their lands and so seek to kill them off. In general, fear and dread of the possible consequences are bred in conflicts of culture complexes; the trait complexes favoring such responses may be wholly objective, but conflicts arise because of the beliefs and customs associated with the traits to be displaced.

The missionary is usually charged with being the arch-instigator of conflicts. Yet many of his teachings need not make great disturbances. Such general concepts as theft, lying, murder, slander, adultery, etc., will be taken by the natives as a matter of course; the differences will lie in the application of these concepts. Thus, adultery may be defined differently by the tribe and the missionary; but usually the missionary comes to teach a new belief, to recount a revelation which exacts new taboos, violates old ones, and so sets up a conflict of beliefs. We usually think of these conflicts as between pagan and Christian, but they occur in respect to any kind of religious propaganda. In our earlier discussion of cults we noted how they might spread from one tribe to another and how they needed a leader and a number of devotees to maintain themselves. Their introduction into a tribal culture will produce more or less conflict according to circumstances. We cite this merely to call attention to the obvious: i.e., Christian missionaries are no different from other missionaries, a gathering host whose origin is as old as culture itself.

The economic school frequently asserts that many conflicts arise from economic situations. What is meant is that the storm centers lie in the traits having to do with food and comfort. It is not for us to evaluate this theory, but to consider economic change resulting from culture contact as provocative of conflict. In 1893 there culminated a religious, militant movement among the Indian tribes west of the Mississippi known in literature as the Ghost Dance religion. A few years before, the herds of bison, upon which these Indians depended for food, were swept from the plains by white hunters. At the same time the tribes were gathered into reservations and refused permission to leave them. Privation was the common experience, and starvation a possibility, though scanty rations were issued by government orders; from the Indian point of view the situation was desperate. At this psychological moment a new religion was launched, a belief in an Indian Messiah, who was to return, sweep away the white race and its culture, and restore former economic conditions. The new belief spread rapidly and revived the hopes of the Indians. By a prompt and energetic concentration of troops, the militant aspect of the movement was repressed, but as a cult or a religion this new trait complex functioned for a long time. The religious movement, the attempt to revolt, may be said to represent a response to a specific economic change. In this case the economic change affected all members of the tribe so that a group conflict was initiated.

The Subject Group

When the group is made subject by conquest, there may be for a time the smart of defeat, but an adjustment will evolve. If, however, the culture differences between the groups are extreme, the dominating group may seem so superior that a feeling of helplessness and despair may prevail. This is perhaps due to a failure to understand the culture to which adjustment must be made. One such group, once visited by the writer, found it incomprehensible why the white official dispensing justice dealt harshly with the native who beat his wife, but turned a deaf ear when a husband complained that his wife beat him. The natives could see nothing consistent in such procedure; they were, of course, wholly ignorant of the historic background to white culture, with its idealization of women. How could the natives understand such a situation, unless familiar with a large sector of white belief? On the other hand, the white practice of demanding virginity of the woman and a lower standard for the male was wholly intelligible to this primitive group, because they followed similar practices. Economic adjustments are also often beyond the comprehension of such a primitive group. For example, an American Indian community was advised that one way to earn some money would be to make hay of the grasses growing in their locality; accordingly they labored successfully during the season, harvesting a large crop. Yet, when it came to finding a market, they failed. White men, they observed, did make a living by producing hay, but not being able to do it, they regarded themselves as inferior, entirely unable to live according to the standards of the new culture. It is this feeling of helplessness in the face of a crisis in culture adjustment that is a significant factor in the conflict problem.

The transition from a conquered group to a dependent one is easy and perhaps natural. When confronted with a crisis for which there is no ready solution, the individual appeals to his fellows; this is true of all, whether civilized or savage. In an analogous manner, a group feeling itself helpless will hope for succor from a stronger group. But under primitive conditions, each small tribe is a law unto itself, must stand upon its own feet; hence, in the face of a culture crisis, they have no group to which to turn for support. Their status is in striking contrast to that of a white colony among primitive tribes, for such a settlement can tap the reserves of a nation in case of need. The natives easily recognize the confidence and security of the white man, and therefore spontaneously lean toward him. Moreover, natives living in harsh environments cannot escape uneasiness as to the future, as instanced by Rasmussen (10) in his studies of the Eskimo. A wise old Eskimo, after pointing out the hungry families in camp waiting for the return of the hunter, the sick and the aged, asked, "Why should this be so?" He said further, "We fear! We fear the elements with which we have to fight in their fury to wrest our food from land and sea. We fear cold and famine in our snow huts. We fear the sickness that is daily to be seen amongst us. Not death, but the suffering. We fear the souls of the dead, of human and animal alike. We fear the spirits of earth and air . . . And for all our angakoqs and their knowledge of hidden things, we yet know so little that we fear everything else."

In his natural state, primitive man usually looks to the supernatural for aid. In more technical terms, he resorts to magic. He feels that somewhere must reside a power able to cope with the danger that confronts him, and that, if he knew how to adjust himself to this power, his problem would be solved. The superior group would be one having such knowledge and skill as would enable it to cope with its environment. To such a group a native tends to turn, as we have stated, but fear may inhibit the impulse. The superiority of the white man is taken to mean superiority in magic as well, a power for evil as well as good. Further, the white man has no understanding or sympathy with the magic of the native, and by ridicule deepens the inferiority feeling of the native. The difficulty, of course, lies in a conflict of culture traits, not in differences in human nature. The native is thus in a conflict of responses on his part, his inclination being to seek the protection and leadership of the more powerful white man, but he hesitates because he fears the white man as hostile, intending to destroy rather than protect. Such states may be attributed to culture conflicts; they are the response complexes of the native group. In most cases of white vs. native, the conflict is wholly one-sided; the culture of the white group is not disturbed by the presence of the native culture.

It may be profitable to examine a few more examples of conflict. The culture patterns for primitive groups seem less individualistic than our own. Their groups are small, but possess great solidarity. The individual is accustomed to act in unison with his fellows; if food is to be gathered, they proceed in a party, etc. Spontaneous response to suggested procedures marks all activities. One phase of the conflict with white culture arises in these differences respecting such a group and individual activities. The outstanding examples of successful native adjustment to economic conditions are cases of adapting the native pattern of group procedure to individualistic conditions. In New Zealand, for example, a few native groups have become self-supporting and prosperous by operating as communities in production and marketing. The point seems to be that whereas a culture may be a conglomeration of traits, it is carried on by a closely knit series of response patterns. It is these patterns that conflict in case of culture contact. The white man expects the natives to take plots of land and scatter in the population as individuals, finding their place in the culture at large. This is asking too much; the social habits of the native follow another pattern; so he tries to operate as a community in phases of culture where the white man operates as an individual. A quotation from Pitt-Rivers (9) is appropriate here. "In all his activity the primitive savage merges his self-regarding sentiments in—or, rather, projects them on to—his social group, either his clan or his tribe. He feels himself an inseparable part of that group, and identifies his passions with those of the other members of the group. There is no sense of sacrificing himself for the good of the whole group. There is merely a spontaneous and unrationalized feeling of identity with the group; just as he regards any member of a social unit outside his own as identified with that outside group."

Finally, another conflict factor may be here commented upon—the tendency of cultures to change. In particular, this is characteristic of white cultures. The adjustment of the native to white cultures must be slow; but the disconcerting aspect of the case is that as the white culture changes, this renders such adjustments as he may have made inadequate. It is difficult enough to bridge the gap between two contrasting cultures, but to see the gap widening as fast as he can build is wholly depressing. In conclusion, then, we see how, on every hand, culture contact between a primitive group and a white presents a maze of conflict that cross-sections the life of the former, resulting in a variety of responses.

Reaction on the Individual

It is a prevailing opinion that when a primitive people come into contact with white culture, they rapidly die off. On the whole, observation does seem to show that during the conflict period the death rate is abnormally high. Various causes have been assigned for this, for one, the introduction of new diseases against which the native lacks resistance or the necessary knowledge to mitigate. This argument has some validity, but on the other hand white settlements in a new country suffer also. The survival of the white settlers is, however, insured by the reserve populations of the home community. The native is further handicapped by the depression resulting from the situations we have just discussed. Usually, the first step in the subjugation of a native group is to remove it to a strange habitat. This habitat may not vary greatly from the former home, but it is strange, and nostalgia results. The suggestion is that anything that seriously disturbs the peace in a native group tends to increase the death rate. Thus in 1838 a large part of the Cherokee Indian nation was forcibly moved from Georgia to Oklahoma; contemporary accounts tell us that about one-half of the cavalcade died on the way and shortly after arrival. The Pawnee were moved from their traditional home in Nebraska in 1874 and settled in their present home in Oklahoma; many died the first year. Further, many young men are reported to have committed suicide, because there were no bison to hunt and wars were forbidden; the future offered nothing to them but to till the soil, to take up women's work. New diseases and malnutrition are certainly to be set down as causes for the high death rate observed, but there is good reason to believe that emotional factors were also operative.

Yet, in the Pawnee example of suicide, we see another aspect of conflict. The new situation created by culture contact leaves the individual no outlet for his aspirations. Where the pattern of individual expression placed social value upon deeds of war and the zest of the hunt, walking behind the plow of the farmer to be rewarded by a few coins made no appeal. The white social values of industry, achievement, and economic independence belonged to a conflicting pattern. A recent study (5) of the Maori in New Zealand shows how the suppression of inter-tribal wars and the elimination of personal combat left the native no outlet for self-expression. His only course was to eke out an existence and live in the mental imagery of the past. However, the New Zealand native seems to have been quite above the average in ability to understand the new culture and to adapt himself. His conflicts were grievous and are still, but on the whole he seems to be conserving much of his old group life and at the same time adjusting to new economic conditions.

In general, then, it appears that culture conflict bears hard upon the individual. It produces a condition bordering upon the psychopathic, analogous to institutional cases in contemporary society. Emotional disturbances are evident. Not only does the group break down in its functioning, but individuals fail. It appears that the severity of conflict is somewhat proportioned to the degree of contrast between the cultures involved.

INTER-GROUP CONFLICTS

So far we have dealt with cultures as wholes; that is, we have sought to observe what takes place when two cultures come into collision. On the other hand, we find that the adaptation of one culture to another is often attended by conflicts within the group. The ethical values of white culture are perhaps best presented in the culture complex which we call Christianity. One of the great organized drives against a primitive group is seen in the process of Christianization. This proceeds by seeking individual converts. One prominent characteristic of new converts to any cause is their hostility to the beliefs they have thrown over, and as soon as a reasonable minority of the native group takes up the new teaching, they form a bloc in opposition. Such a conflict may become a serious menace, particularly if the minority has the backing of the white group. We meet with this in the United States, when the Christian Indians on a reservation seek the aid of the government in suppressing social gatherings and ceremonies among the opposition. In objective terms, the phenomenon may be stated as an inter-tribal conflict of social values. One observed result in such situations is for the conservative wing to draw more into isolation, to reduce contact with outsiders to the minimum, to resist suggestions from without, in short to induce a state of mental and social stagnation. Yet this type of conflict in less extreme form, seems to operate in all culture groups and may in large part be responsible for the constant change we noted in a previous section. So-called reforms are brought about in this way. In our discussion on borrowing, we noted that the fate of an introduced cult hung upon the skill and faith of its initial devotees. The traditions of primitive folk often herald the successes of the innovator in introducing new values. One American tribe treasures a narrative of how a young chief at the risk of his life abolished the custom of human sacrifice. When such conflicts arise in a tribe leading a free natural existence, they usually result in separation and the formation of a new tribe; both may then proceed with a minimum of internal conflict. On the other hand, a tribe held in subjection cannot solve the problem so easily and consequently may suffer an intensification of inter-tribal conflicts. Something like this is observable in class conflicts in national populations, but these are fully discussed in works on sociology.

Survival

To complete this review of the subject, it may be well to consider the factors that make for the survival of cultures. The group feels uneasy if anything threatens to take its culture away. The emotional and other responses we have noted in specific cases of culture conflict are most intense when the culture breaks down. So the normal functioning of the community depends largely upon the smooth operation of its culture, and the survival of mankind indicates the presence of protective responses. In sociology, comment is made upon tradition and resistance to change as a detriment in so far as progress may be inhibited, but as an advantage in stabilizing culture or in controlling innovations. The reactionary response, when viewed philosophically, is one of the protective factors in guaranteeing the survival of cultures. We have noted that cultures change spontaneously without jarring the group and so without invoking any of the protective responses. On the other hand, an abrupt change, an innovation, will invoke these responses.

One aspect of culture not adequately explained by social science is the balance that seems to exist among the many traits composing it. It appears as a finely balanced whole; if the economic basis drops out, the whole structure threatens to go down. Yet, in much the same way, the breaking-down of a taboo system may topple over the whole structure. The destruction of established values, as head-hunting, scalp-taking, etc., has also been observed to disarrange the whole culture. At first this may seem inconsistent with our previous discussion in which cases were cited of the taking-over of new traits successfully, but cultures grow in large part by reacting to each other in this way. It is only under certain conditions that the balance holding cultures in function is disrupted. Since such disruption follows white domination, it behooves us to learn to avoid the conditions that favor such upsets. One set of problems in economics and politics deals with the maintenance of balances between cultures.

Pitt-Rivers (9) has attempted to explain declining birth rate and vitality as due to emotional states resulting from culture conflict. He thinks there is a delicate balance between the biological functioning of the organism and culture; that, if there is a real conflict, the resulting emotions act upon sex and nutritive functions to such a degree as to disturb organic life. The inference from this is that there is a fine adjustment between culture and the human organism. If this is true, we may expect culture conflict to be accompanied by emotional disturbances.

The question is often discussed as to whether isolation favors the survival of cultures over long periods with little change. We have noted that culture contact under normal conditions favors progressive change, but have not considered the behavior of the group in isolation. It is usually assumed that the cause for the backward state of such peoples as the Tasmanians, Andamans, and Fuegians was their lack of contact with other cultures, but we can best approach the question through examples of adjustment in partial isolation. In the course of immigration into the United States, numerous communities of foreign folk have been established, some of which have survived over long periods. A study of a few such show that they have survived by making the minimum economic adjustment, but otherwise avoiding all the culture contacts possible. The core of organization is religious and social. The peculiarity, however, is that in such communities many aspects of the culture are about what they were in the home land when the group emigrated, so revealing themselves the more conservative of the two. In these cases, the community sought by isolation to resist change and so approached a stabilized level. Such survival may be said to be largely self-determined, in contrast to the primitive groups cited above. Parallel cases can be found in subject peoples, as among some village Indians in southwestern United States. These Indians were agricultural when discovered and have continued self-supporting to this day; thus their economic adjustment is satisfactory. On the other hand, they have drawn aloof from white contact; fearing ridicule, they have carried on in secret, putting forth every effort to conserve things as they were centuries ago. In this they have succeeded to a large degree. However, these examples of survival by withdrawing from culture contact rather strengthen the case for the growth of culture through the interaction of culture groups.

In conclusion, we have reviewed a phase of psychological inquiry in the making. No great leads to research in this field have been discovered, but, since the other chapters in this book are eloquent pleas for the efficacy of experimental approaches, it seems a reasonable expectation that success in this field will come with the application of similar methods. While humanitarian ideals may make direct experiment impossible, governments are setting experiments for us. The statements made in the preceding pages are based upon observation on the behavior of groups under specific conditions. Such studies can be carried out systematically, are in fact now being initiated. On the other hand, some group experiments are possible, among children, communities, etc., and may be expected when the observation studies of selected culture groups have progressed far enough to make the formulation of test experiments desirable.

REFERENCES

1. ALLPORT, F. Social psychology. Boston: Houghton Mifflin, 1924.
2. BARTLETT, F. C. Psychology and primitive culture. Cambridge: Cambridge Univ. Press, 1923.
3. DEWEY, J. Human nature and conduct. New York: Holt, 1922.
4. FRAZER, J. G. Totemism. Edinburgh: Blank, 1887.

5. Keesing, F. M. The changing Maori. New Plymouth, New Zealand: Board of Maori Ethnological Research, 1928.
6. Malinowski, B. Sex and repression in savage society. London: Kegan Paul, 1927.
7. Morgan, L. H. Ancient society. Chicago: Kerr.
8. Ogburn, W. F. Social change with respect to culture and original nature. New York: Huebsch, 1922.
9. Pitt-Rivers, G. H. L. F. The clash of cultures and the contact of races. London: Routledge, 1927.
10. Rasmussen, K. Across Arctic America. New York: Putnam's, 1927.
11. Ross, E. A. Social psychology. New York: Macmillan, 1912.
12. Sprowls, J. W. Social psychology interpreted. Baltimore: Williams & Wilkins, 1927.
13. Tylor, E. B. Primitive culture. London: Murray, 1920.
14. Wissler, C. Man and culture. New York: Crowell, 1923.
15. Young, K. Source-book for social psychology. New York: Knopf, 1927.

CHAPTER 21

THE ABNORMAL INDIVIDUAL

SHEPHERD IVORY FRANZ
University of California at Los Angeles

The individual is not only a development from united germ-plasms but he is also, and some say primarily, the result of various external forces acting upon that germ-plasm combination. Peculiarities of the developed individual which distinguish him from others may, therefore, be the results of hereditary influences or of environmental accidents. This is more obviously the case with the abnormal person. Born from a stock which we say has not developed cerebrally, he may also show less cerebral and mental development than most of the community in which he lives, and we call him feeble-minded. Or, he may have developed from a combination of elements from male and female who are in most respects like their fellows, but because in childhood he received a blow on the head or because of a bacterial invasion, as in meningitis, he may exhibit differences in behavior which make him very different from his parents, and from his father's and his mother's friends and acquaintances. Thirdly, he may have inherited what is commonly called a "nervous constitution," possibly a combination of peculiarities of his parents, and because of that constitution the ordinary events or incidents of daily life give rise to actions or ideas markedly peculiar to his age and to his social station. On the other hand, hereditary and environmental effects may be combined in such a way that a person develops quite differently from his parents and their environmental friends, and although each individual action may not appear peculiar, the combinations of his actions are distinctly unusual.

The abnormal individual is essentially a peculiar or an unusual person. He differs from other people in many ways, but when we analyze the differences we find they are mainly, and perhaps wholly, those of degree. If one pictures to himself the characteristics, physical and mental, which he is accustomed to find in the carpenter, the professor, the street-sweeper, the shop-keeper, the physician, the bond salesman, the conjurer, and a host of others, he discovers that while he may recognize a diversity in many respects he also sees that they have certain general resemblances. Some of the observable differences may be characteristic for occupations, but many may not be characteristic of them. It is well known that clergymen may be either tall or short, laborers may be talkative or taciturn, the actor may swagger or be unnoticeable in a crowd by any behavior mark, and so forth. Although there are wide divergencies in most social groups, individually both in physical make-up and in behavior, the differences are not great enough to cause many persons to be thought of as outstandingly peculiar or unusual. Those that do stand out as especially unusual, or as great deviates, make up the abnormal group. They may be abnormal in either direction. The renowned lawyer may be as great a deviation from the crowd as the stupid laborer who exhibits no initiative and who must be instructed each time in each operation how he is to carry it out.

There are four large groups of mentally abnormal people. These are respectively: (*a*) the feeble-minded, (*b*) the psychotic, (*c*) the neurotic, and (*d*) the eminent or the geniuses. The feeble-minded are mainly those with underdeveloped brains or those in whom cerebral and mental maturation is slow. The psychotic are those in whom the brain and the mentality dependent thereon have developed to a certain degree, and in many cases have matured in a normal fashion, but who because of accident have become mentally twisted. Some of them have regressed so that they are not only peculiar for the group but they are also recognized to be unusual in relation to their previously exhibited mental states. The neurotic are those who are frequently described as having an "unstable" mentality. Against them slight internal and external events work bringing about relatively great deviations from the behavior of the average man and from the individual's own "normal" mentality. The eminent are those who deviate from the average in an excess of mental accomplishment. This excess may not be a matter of quantity of performance but one of quality, as judged by those in the same line of endeavor.

In general, the feeble-minded are those that develop and mature slowly, but in this group there are classes, artificially determined as to limits, to which separate names have been given: idiots or aments, imbeciles, and morons.

The idiot is essentially a person who exhibits a condition of limited and slight mental development. The brain, especially at the frontal pole, does not develop to the extent that is found in most children, even after many years, and corresponding or concomitant with the failure in brain growth there is a retardation in mental growth and status. In some instances the stimulus-response relation is practically at the reflex level, so that the child (regardless of his age) responds only to a few simple sensory stimuli by simple responses. In neuro-mental development he may be at the level of many nervously simple organisms, in which direct responses are obtainable but in which many stimuli do not give rise to indirect responses. Food may be taken as does an animal, but the idiot may not be able to learn to use the indirect means of feeding himself with knife, fork, and spoon. Clothes may be used, but rather as a general covering like leaves than as garments which must be put on in special ways and which must be buttoned or hooked. The social characters of normal maturing adolescents and of matured individuals are non-existent. Cries, grunts, and other inarticulate sounds may be produced, but these vocal sounds are associated directly and solely with degrees of satisfaction, discomfort, pleasure, or pain. In some instances mental development progresses to the point of some slight social activity such as expressing need for bodily attention. Generally, however, there is lacking the social adaptation to the degree of ability to feed himself, to care for his other bodily needs, to express understandingly to others his wants and his pleasures, and to carry out those social accomplishments of dressing, talking (and, of course, reading and writing), preparing for the morrow. In some instances the lack of cerebral and mental development may primarily be due to, dependent upon, or concomitant with, developmental retardations of structures other than the brain. Not infrequently we encounter those who lack development of the thyroid gland. The function of that gland is one concerning general metabolism and general bodily development. To those children in whom general mental-cerebral relation is retarded because of failure of development of the thyroid gland, the name of cretin is given. This is due to its early recognition as a frequent occurrence in the island of Crete. Regardless of the number of years he may live, the idiot usually does not attain the intellectual level of the average three-year-old child. In other words he cannot carry on those mental—and social—operations which are characteristic in a child of three years of age. It is this dividing line which is taken to distinguish this slightly developed group from the next higher group.

The imbecile grows to the level of the seven-year-child and he may vary from the three-year-level to this higher point. He learns to care for his bodily needs, he dresses himself, his vocabulary is sufficient to indicate his needs, his desires, and his interests. He may learn to do many relatively simple operations of an indirect character, such as driving a truck or caring for a garden, and he may learn to write his name and to "read" the pictures in a comic strip, but he usually fails to get the significance of more than those numbers below 20, and he does not read anything but the most direct and the simplest statements. His social activities are limited, although he may be apparently polite and submissive to conventions. When sufficient educational training is available, he may become an automaton worker in the social world, who reacts to individual stimuli, mostly simple but some relatively complex, in a direct fashion. His reactions are seldom indirect. The stimulus of the present moment is not often effective for reactions of tomorrow. In many cases each task is a thing in itself. He may, however, be taught to carry out a number of tasks in sequence if they are related. In many homes and institutions imbeciles do most of the routine work, mainly under direction. If they have been properly instructed, they do well what they have to do. They are like machines, adapted to a few tasks which they carry out in a regular fashion once the machine has been started. Many of them must, however, be told when to start, when to stop, when to change, when to omit some simple thing. They do not "reason" why, they continue to do the job before them as they have learned to do it. They can learn to lay bricks, or to drive nails, but they cannot learn to build a house. Opposed to these limitations there may be a development of special aptitudes of a remarkable character. One may have learned, in some unknown manner, to make mentally arithmetical multiplications, even of four- or six-place numbers, rapidly and accurately. Another may have acquired, also in an unknown manner, the art of determining and telling the day of the week on which any monthly and yearly date came, as, for example, January 2, 1922. Special aptitudes of many kinds have been described for other *idiots savants*, as they are called, but these phenomenal performances are associated with the usual imbecile inadequacies in other directions.

The moron is the main "hewer of wood and drawer of water" in the industrial world. He who called himself a machinist but who described his work as "putting the nut on the seventh bolt" of a manufactured article is typical of this class. They are more intelligent than the imbeciles; they have some initiative corresponding to the level of the 10-year-old child; they know what to do from day to day; they can adapt to many more, and to some fairly complex circumstances; they may take part in general social community activities, but they are not initiative; they may read the newspapers, especially the sport section and those parts relative to local affairs; they are interested in the things of the moment, but know little of the past and they plan only superficially for the future. The mechanical operations of the shop or the factory, the routine in the industrial or commercial world, and the humdrum of the movies or the theater indicate their level of attainment.

Except for the general retardation in development there may be no structural or functional abnormalities in the idiot, the imbecile, or the moron. Many idiots do exhibit both anatomical and physiological disturbances, some imbeciles exhibit corresponding conditions, but the moron is as free from these as the so-called normal. The mental retardation in the imbecile or in the moron is not, therefore, due either to a motor irregularity or defect which prevents normal behavior, nor is it due to the lack of ability to take things in through the eye, the ear, the nose and mouth, or the skin. There can be no doubt that in some instances the motor or the sensory endowment may be somewhat limited, but there is reasonable doubt that these defects are primarily the cause of the mental retardation. The favorable results of educational efforts with some children born blind or deaf show that the sensory defects in themselves do not give rise to feeble-mindedness. Similar results with children who lack coordination because of neurological defects or disease (such as early destruction of portions of the cerebral cortex, as in birth paralysis or in encephalitis) are evidences leading to the same conclusion. When, however, there are sensory or motor defects associated with the general neuromental developmental retardation, the result is less favorable to learning and social adaptation. Feeble-mindedness is, therefore, not necessarily the result of primary disturbances in the receptive or emissive spheres, but it is probably the effect or an expression of a neurological state in which associative linkages are either not anatomically developed or not functionally workable. This belief is supported in part by investigations of Bolton and others of the brains of idiots and low-grade imbeciles in which there was observed a lack of development of at least one of the cerebral association areas, viz., the frontal.

That other developmental defects may give rise to retarded mentality has been shown by the observations of a special class of feeble-minded whom we have called cretins. These children do not develop physically as do normal children. They retain infantile characteristics for a relatively long period. Anatomical studies have demonstrated that the children who make up this class are deficient not only in general development of bones, muscles, brain, etc., but especially in the absence of the thyroid gland. When, however, in these cases the child is fed with the extract of the thyroid gland, there is a progressive but gradual improvement both in physical status and in mental development. If the medicinal treatment is continued for a sufficient period of time, the affected child develops a normal physical condition and a normal mental state. Not infrequently improvement in the mental as well as the physical condition is noted immediately the gland extract is administered. As time goes on, it is noticeable that the beginning of development generally corresponds to the date of administration of the thyroid extract. If, for example, the nature of the defective development is recognized at the age of two, and the thyroid administration begins at that time, at the age of six the mental age which is attained may be about four, and at the age of eight it may be six years. In other words, the "normal" development begins when the deficiency in thyroid secretion is compensated by thyroid administration. Some retardation may be apparent, but developmental possibilities are also obvious. If treatment continues, there may be an apparent mental retardation, but there is a correspondence with the physical development. The child of eight may have a mentality of six years and a corresponding physical development of six years. At 12 he may also be retarded in both physical and mental states by two years. If treatment is continued, the development may be retarded by the number of years prior to the recognition of the condition, but by the time the patient is 16 years old the mental state cannot be distinguished from that of the child who has always been normal.

Recent investigations of retarded or psychotic individuals have shown that, although indications of glandular disturbances are not so exact as those known to be associated with the thyroid gland, other glands may be equally important for development and growth. To some imbeciles the administration of extract of the pineal gland appears to have brought about improvement. The clinical determination of the cases to which this therapeutic administration is applicable is very much in doubt. There seems, however, reason to believe that there is a limited number of imbeciles and perhaps also of idiots, in whom there is faulty development of the pineal body and to whom the administration of the extract of that gland may be beneficial.

These cases of retarded mental and physical development which are the the concomitants or the results of glandular insufficiency are paralleled by those in which accidental conditions alter the endocrine secretions in individuals who have developed. In some developed individuals there are mental abnormalities accompanying decrease or lack of some glandular secretion, but in other cases mental abnormalities are associated with increased secretion of some of the glands.

Mental disturbances which are associated with increase or decrease in the activity of the thyroid gland have been most carefully investigated. When there is an increase in thyroid secretion (the condition may be produced if there is administered thyroid extract to a normal person), the individual exhibits signs of general "nervousness." He may have tremors of lips, tongue, and fingers; his heart beats more quickly; he is irritable and reacts quickly; he is emotional and very frequently fearful; he may be flighty so that he changes from one thing to another; he may want to keep going when under normal conditions he would become fatigued; with these changed conditions he may lose considerable weight. The opposite condition of decreased thyroid activity results in increased weight and bodily coarseness; the mental functions are sluggish, and the patient does not want to attend to things which are necessary; his emotional reactions are slight and few; he tires easily.

The mental changes due to thyroid dysfunction give warrant for the suspicion that mental conditions are also modified by alterations in the functions of other glands. Many of the relationships have not been determined, but there are some characteristic mental conditions associated with functional abnormalities of the pituitary, the adrenals, and the gonads. Decrease of adrenal secretion is accompanied by a lack of vigor, both physical and mental. We know little about the "normal" effects of adrenal activity, but in emotional states the adrenals are known to be more active and their increased secretions give rise to disturbances in the heartbeat, the caliber of the blood-vessels, the secretory organs, the pupils, the skeletal muscles, and the gastro-intestinal system.

Because of the obvious changes in mental states due to increased or decreased function of some of the glands, there has been much theorizing regarding the actions of all the glands in the production of those mental derangements which are grouped under the general name insanity. Many years ago Kraepelin (22), speculating upon one of the most common mental diseases, dementia precox, suggested that it might be due to thyroid malfunction. A recent modification of this view is now noticeable in the attempts at correlation of many supposed anatomical, and necessarily functional, developmental anomalies with the mental states in dementia precox and in other psychoses. When, however, these attempts are examined statistically, they do not pass the criteria of the simplest procedures. All of the views are extensions of the view that every mental disease is due to or is concomitant with disease of the brain. That view in turn is based on the hypothesis that all mental states are dependent upon brain activities. Explanations of this kind have been designated by Southard as "brain-spot" hypotheses of mental disease.

In apparent opposition to these views are others which Southard has called the "mind-twist" hypotheses. The latter, exhibited in its extreme form in the explanations of the psychoanalytic schools, are being extended to cover the mental symptoms in the so-called organic psychoses in which for many years cerebral changes have been known to be accompaniments of the psychical abnormalities.

In most instances those who adhere to one or other of these explanations fail to realize the philosophical import of their hypotheses. Those who hold to the "brain-spot" views are dualists or materialistic monists. Those who advocate the "mind-twist" hypothesis are philosophical idealists. From both psychiatric and psychological viewpoints it should be recognized that description in our present state of ignorance regarding most of the underlying causes and mechanisms may most adequately be made if we hold to neither view systematically, but describe the conditions as best we can, shifting from the brain-spot to the mind-twist terminology, and vice versa. If the criticism is urged that this does not lead to a unified view of mental derangements, the reply is that the hypotheses must bring together the facts, not that the facts should be suppressed or distorted to fit the hypotheses. This, it will be noted, is what we have done above in connection with the consideration of the feeble-minded. Some of the developmental retardations are obviously associated with anatomical and functional abnormalities of the brain and other organs; others are at this time describable not in relation to structural peculiarities but solely as conditions of slow and limited learning.

It has been usual to divide the mental diseases into two classes—those in which there are obvious pathological changes in the brain and those in which the mental abnormalities are not associated with cerebral disease. In two different individuals, as well as in two different diseases, some of the mental symptoms may be the same. Hallucinations, delusions, states of depression or of exaltation, and many other symptoms, are to be found in patients with marked cerebral changes as well as in those whose brains exhibit no deviations from what we consider the normal cerebral state. On the other hand, it is known that two patients may exhibit in general the same symptoms but their brains show dissimilar destructions. Conversely, two patients may exhibit different symptoms but their brains show apparently similar destructions. If, on the other hand, we consider a disease for which there is no known concomitant cerebral change, we find that in two patients the symptoms are not the same. There may be similarities in the main, but in the elements they are markedly different. This will be more noticeable as we proceed, because it will be seen that in special diseases patients may or may not exhibit the same symptoms. In one disease, manic-depressive insanity, it will be noted that the symptoms may appear to be entirely different in two patients.

It is customary to begin the discussion of mental diseases, commonly known as psychoses or insanity, with the consideration of the mania-melancholia condition (manic-depressive psychosis or cyclothymia), or with the dementia which begins at or about the time of puberty and which progresses at times to the point where the patient is apparently the mental equivalent of a vegetable (dementia precox), or, thirdly, with a mental disease due to an infection which attacks the brain and which results in the loss of mentality and of physical power (viz., general paralysis of the insane). Here, however, it seems more convenient and more appropriate to consider first those mental changes incident to old age.

While there is general recognition of the fact that with advancing age there is a lessening, and at times finally a loss, of mental vigor and grasp, the reason for this is often not understood. We see some people who, at 90, are the mental superiors of others at 60. When, however, we begin to examine more carefully the brains of these persons who may have died of intercurrent diseases, we see there are great anatomical differences. In the case of the feeble-minded we noted that at the age of 20, one child may show abilities of the child of four years, and another may show abilities corresponding to the average child of eight years. So also we note anatomically that the brain of a man 40 years of age may closely resemble those of others who are 60 years old, while the brains of many who are 80 years of age may also correspond in structure to the brain of the individual who is only 40 years old. When many brains have been examined, we find that in some the arteries have been partially blocked by accretions which harden them and which prevent the free and full passage of blood. In these cases where arteries have been partially blocked the brain areas supplied by those arteries are not well nourished, and they decrease in size (atrophy). The partial blocking of blood supply, with the resultant decrease in nourishment, is the usual condition in old age. The blocking is, however, relative. When it is complete, or almost complete, a part of the brain may atrophy so that it almost disappears. When, on the other hand, the change in the arteries is relatively slight, there may be a slight degeneration of the brain, but in many instances the tendency to blocking the blood supply is compensated by increased heart action. When the blocking is completely compensated, no brain change takes place.

In chronological old age there may be no nutrient block owing to compensating heart action, but in mental old age there is a well-recognized block. In mental old age the brain atrophy, which is progressive, gives rise to or is concomitant with characteristic symptoms. The most noticeable and the most common of the symptoms is disturbance of memory. Occurrences of the moment do not register. They may be taken in, as can readily be demonstrated, but they leave no effect. If the patient be stimulated strongly, he may respond in normal fashion. If, however, he is given a stimulus to which he responds correctly, and in a day or in an hour he is questioned about it, he most frequently is certain that it did not occur. In almost direct opposition to this (from the mental standpoint) is the ability to remember the events of early childhood. While this is taken as a characteristic of the senile, or second-childhood, stage of mental life, one should remember the investigations of Galton (31, 32), who discovered that in adults many associations are those from early childhood. These recalled associations probably owe their dominance to the relative simplicity of mental life in the child, and to the corresponding values of the individual elements. The senile individual is essentially one who does not take in the stimuli of ordinary intensity. In order that he shall react to the stimuli which come to him, the stimuli must be intense or prolonged. When attention is called to special happenings, and the patient is compelled to attend, he manages to grasp them so that he is able to identify them at a subsequent time.

Because of his inability to recall what happened an hour or a day previously, a patient of this kind will sometimes attempt to fill the memory gaps with "normal" events. If asked whether or not he knows the questioner, he may peer at him intently and then say that he thinks his visitor is some person with whom he was formerly intimate. If asked where he was the day before or ten days previously, he may answer glibly that he was at his office, or that he was playing golf, or that he was doing something else consistent with his previous life. If pressed further, he will recount incidents of his occupation at these times, the people whom he met, the golf score he and his opponent made, and other items, smaller and more intimate. All of the incidents cited may, however, be untrue—manufactured or imagined by the patient. They are imagined because he has a lacuna, a vacancy, in his memory, which he attempts to fill in. The items he uses as fillers are those which are of the greatest interest to him—those which he "wished" had happened. More careful investigation shows that the patient is often willing to admit that he is somewhat uncertain, and at times suggestions given by his visitor are readily accepted. If they are not accepted, because they are rather generally opposed to his accounts, they put him in a doubtful mood. On the following day, however, he may identify wrongly the same visitor, and tell him the same tale of having been fishing, or golfing, or attending to his business. In many instances the memory lapses are filled in day by day with the same supposed occurrences.

From a psychological point of view the two elements of most interest are that the patient does not take in or does not retain the effects of occurrences as does a normal person, but that he pretends to do so. The philosopher, John Locke, would probably have said for the first that the wax tablet of memory was either too hard for the stimuli to leave much impress, or that it was too soft so that the lines are easily obliterated. An analogy of this kind does not make clear the occurrences of substitution, and the Freudian explanation that the substitution is a "wish-fulfilment" is not much better.

In this connection it is of interest to note that the application of mental tests to patients of this character does not furnish results of any value. Intelligence quotients arrived at in the usual manner are worthless. They depend almost entirely upon the examiner's efforts, and they do not represent the condition of the patient who is examined. They appear to coincide with the old belief that the senile individual reverts to a childlike condition. The tests fail scientifically in that when the stimuli are intense they often give rise to accurate responses in the senile state. This is different from the conditions in children.

Closely allied to the senile state are those conditions in which the brain, because of arterial disturbances due to poisons, such as alcohol or those accompanying bodily waste, becomes affected in a secondary manner. The arteries may become encrusted with lime deposits which prevent the normal flow of blood, and various parts of the brain become atrophied or reduced in size. This obvious change is the gross state which is microscopically evident in great changes in the nerve-cells.

In these conditions the mental picture varies in accordance with the location and the extent of the atrophied portions of the brain. When those parts concerned in movement control are affected, there may be great weakness, sometimes with tremors, and there is the tendency to inactivity. Sensory disturbances are not uncommon—visual, auditory, olfactory, gustatory, and cutaneous. Accompanying the sensory disturbances there may be hallucinations. There may be mental irritability or a state which appears to be one of indifference. Part of these is probably due to the fact that the patient, like the senile person, fails to take in situations, and his memory is poor. Because of these, he reacts poorly in new surroundings, is not adaptable, and most frequently shows poor judgment in ordinary affairs of life. The fact that he is not bothered because of his poverty, his failing business, or his associates is indication of his inability to sense social situations.

These degenerative or demented conditions which are due to long-continued and excessive use of alcohol are not the only mental disturbances which this poison produces. Acute, transitory, mental symptoms may be produced by over-indulgence over a comparatively long period, followed by a period of abstinence. It seems as if the nervous system becomes accustomed to the alcoholic stimulation, and when the alcohol is withdrawn, disturbances of the nervous system result. The most common state of this kind is called delirium tremens. The name characterizes the condition well, in that the most noticeable feature is that resembling a delirium. There are hallucinations, mostly of vision and the tactile senses, some of which give rise to ideas of a horrifying character. They may be of a persecutory type, or of facing death by the trampling of green elephants, or being crushed by horrible red snakes, or bitten by huge spiders, and the like. At times the patient imagines he is carrying on his usual occupation although he may be confined to bed. Along with the delirium there are tremors, which make movements uncertain and irregular. Notwithstanding the motor derangement, the patient may be continuously active, thrashing around in bed, pacing the floor, hammering upon the walls, shouting, or making efforts to escape his imaginary human or animal persecutors.

Notwithstanding the distinctively morbid states which characterize patients of this kind, they may be "waked up" for a few moments by suitable verbal stimuli and respond rationally, but soon revert to the state of delirium. Control of the delirious symptoms may be brought about by the use of alcohol or by the administration of sedatives or hypnotics. The latter tend to dull both the brain and the sense-organs and the decreased irritability lessens the tendency to sensory disturbances. At the same time they reduce the motor excitability.

Somewhat similar symptoms of an hallucinatory character follow the use of certain other poisons or drugs, at times following the administration of the drug in relatively large quantities, and at times in an addict when the drug is taken away completely and suddenly. The most common conditions of this kind are those which follow the use of opium or one of its derivatives, morphine in particular. Similar states follow the use of cocaine, hashish, and mescal. All are characterized by hallucinations, some of which are pleasant while others are unpleasant. In addition there are usually, because of the blunting of the nervous system, difficulties in taking in situations and of reacting to them appropriately. This gives rise to many unsocial acts. In many instances the patient has a fundamental "suspicious" state, which colors his experiences and which causes him to assault his wife, to abuse his children, or to try to harm his employers. The special drug acts on different parts of the nervous system differently, and it is partly because of this that there are differences in the hallucinations and in the delusional conditions.

Because of the recognition of well-defined cerebral changes and because of the attendant belief that the mental diseases were dependent upon cerebral destructions, general paralysis of the insane was for many years considered the paradigm for all psychoses. At present we recognize this disease to be a particular form of infection or toxemia, in which the infectious agent is known to be the spirocheta pallida, the organism producing syphilis. Several, formerly supposed to be relatively independent, mental diseases are now known to be forms of the same general disease, differing because of the character of the invasion of the pathogenic organism. In some instances the brain is affected soon after the occurrence of the infection, whereas in other cases the mental symptoms do not appear to develop for several years after the initial general bodily symptoms of the disease. There has been much speculation regarding the reasons for these differences, only two of which need mention. One is that in different individuals there are different degrees of immunity against the infection, or, conversely, degrees of chemical affinity of the nervous system and the spirocheta. The second is that there are different forms of spirocheta which may have chemical affinities to attack different parts of the nervous system. These, it should be noted, are speculations.

What we know in general is that in some patients the onset of the mental disease is sudden and the symptoms resemble an acute poisoning or infection of the brain, whereas in other patients the onset is insidious and the general symptoms more nearly resemble those found in old age or in arterio-sclerotic brain changes. In both forms, and there are intermediate ones, there are differences in symptomatology which are fairly distinctive. Some are general physical signs indicating involvement of parts of the nervous system other than the brain, others are indicative of disturbances of well defined brain parts.

In some cases of syphilitic infection of the nervous system, especially in the early cases, the mental symptoms are general, including headache, disturbances of the sense-organs, and slight motor difficulties. If at this stage efficient treatment is instituted the symptoms may disappear and the disease be terminated. Some patients do not develop mental symptoms soon after the infection, but these symptoms may be delayed for three or four years, at which time there is a rapidly developing mental disturbance, which in addition to the symptoms enumerated above may include a comotose-like state, in which the patient pays little or no attention to his surroundings. Mott has compared the state to a semi-intoxicated condition, in which there may be delirium, from which the patient may be aroused for a short interval by suitable stimuli. This mental disturbance may pass away, leaving the patient in an apparently normal mental state, to be followed later by another and perhaps prolonged similar series of symptoms. This form of syphilis resembles both in its pathological attack and in its symptomatology other mental diseases which are due to infection of the brain coverings (the dura mater and the pia mater).

In other instances the syphilitic germs apparently attack the arteries, producing in them conditions that correspond to premature senile or arterio-sclerotic changes, preventing proper nourishment of cerebral parts and also bringing about abnormal excitation of parts. The symptoms are essentially those which have been described above under arterio-sclerotic mental states.

Another chronic syphilitic brain state is that which was formerly known as general paralysis of the insane. In this disease, there are many motor and sensory disturbances, which are accompanied by what is supposedly typical of the disease, viz., expansive, foolish delusions. The motor disturbances are usually those of muscular weakness and of irregularity of movement. The weakness causes the patient to remain in bed, or if he walks about he is apt to fall and injure himself. The irregularities of movement may exhibit themselves in the form of tremors which cause vocal speech to be thick and difficult to understand, written speech to be irregular and at times a mere scrawl, and other movements are so incoordinate that the patient is unable to care for himself, in eating, dressing, or in any of the daily rounds of life. The senses (from the periphery to the central stations) may be so obtunded that the patient does not "see" important things to which he should attend, and he may walk about in a naked state without feeling the cold. Because he does not take in the items surrounding him, he may not know his friends (because he cannot identify their features) or his home (because he cannot differentiate it from other houses) or the time of day or of the seasons. His ability to attend to events may be sharpened by intense stimulation, but it is usually low. Because of these things he has poor judgment in practically every affair, and he is guided largely by his reflexes (or instincts) which may also be disturbed.

Although the supposedly characteristic form of delusion is the expansive, grandiose type, there are some patients who are sad or who have neither melancholic nor maniacal delusions. Practically all of the delusions are ill balanced and absurd. If the patient is grandiose, he owns the world, or has ten billion slaves, or is the greatest inventor (although he may not be able to tell what he invented), etc. All of these ideas are voiced even in the conditions of lack of clothes or food. On the other hand, he may tell how he dies every day (but he eats his meals and meets his fellow patients in a casual manner), how he is always being tortured and that his legs have been pulled off, that his body is gone, or that his internal organs have been taken away for other people and have been replaced by wooden or metal organs, etc.

Along with the bizarre delusions, which may be constant but not systematized, there are definite memory disturbances.

The physical disease of the brain often leads to convulsive seizures, in one of which the patient may suddenly die, or there may be a condition of inanition (bodily starvation) due to the degeneration of all the bodily functions.

The mental deterioration in those cases which are definitely known to have destructions or atrophies of parts of the brain is paralleled by a corresponding but somewhat different kind of deterioration. This is a dementia in patients who are young in years and who have not brain changes which many psychiatrists are willing to admit as causative agents or as necessary accompaniments. In this dementia the symptoms of the mental diseases are exhibited sometimes as early as the 15th year, and about three-quarters of the patients of this class admitted to hospitals are less than 40 years of age. Because of the early or precocious development of symptoms and because in the great majority of cases of this type there is an eventual profound dementia, the disease has been called dementia precox (or praecox). Some investigators have found microscopical changes in the brain in some patients who have had this disease, and both Southard and Rosanoff (9, 27) have discovered relatively gross cerebral lesions, viz., evidences of atrophy, similar in some respects to those in other dementias. On the other hand, it should be said that Southard has failed to find brain changes in about half of the cases. Although there may be differences in causation there has been here, as in relation to some of the mental diseases to which we have previously referred, the tendency to attempt a single and inclusive anatomical-mental or a physiological-mental correlation. Among the hypotheses there may be recalled the early one of Kraepelin (22) that the disease had definite relations to thyroid malfunction. Subsequently there have been extensions and elaborations of this early hypothesis, but especially noteworthy are those explanations which predicate deficiencies in development of anatomical structures and physiological activities. Among these explanations there may be mentioned, but only as speculations of possibilities the references to the disturbances in function of the gonads (or sex organs) or of the circulatory system, and to the disturbances in many organs which may be localized (as examples, in the tonsils, in the teeth, or in the uterus), but which because of toxic effects bring about general effects in many other organs and especially in the nervous system, which is conceded to be the "organ of the mind."

However, because of the failure to demonstrate a unitary anatomical relationship or a physiological concomitant of the mental disturbance, there has been considerable discussion of the probability of a peculiar cerebral organization or a special mental substrate for the disease. Among those who have discussed this matter is August Hoch (35), who suggested that there was a particular mental make-up or personality in most of those who subsequently develop dementia precox. This personality is generally described as the "shut-in" type. In early years the child may have been considered quite normal. He may have learned readily in school and kept up with his companions intellectually, but he may also have taken little interest in the plays and games at school and in the home neighborhood. He, perhaps, was not unsocial, for he was pleasant, but he was rather quiet, with a tendency to studiousness or to daydreaming. Later this developed into an aloofness, and he reached a degree of seclusiveness which was noticeable. Then a mild degree of the special characters of the dementia precox became apparent. There was not only seclusiveness, but inactivity; daydreaming might have become an almost constant state, and when the patient was interfered with because he was thought to be lazy he may have resented it. He may have realized that he was not getting on in the world, but he consulted no one about his difficulties. He thought he was especially apt to be misunderstood or he was actually misunderstood, and this led to further difficulties. At times he became quite unsocial; he might, for example, have appeared to be quarrelsome. At other times he might have attempted to escape from the difficult situations by wandering from home. Again, there were obvious illusory states, in which the situations appeared unusual or even unreal, and these unreality situations gave rise to peculiar reactions. Occasionally there were hallucinations, mostly visual and auditory, which were not appreciated as dream-like states. They were carried over into the normal waking life, and sometimes as if they had validity of an unusual character.

These special characters in the personality of the individual can readily be understood to be the underlying conditions in the mental disease, especially when they are exaggerated at the time the disease is usually recognized as a well-defined psychosis. At the same time we are able to see that the predominances of certain of the personality traits are the prodromata of the similar but exaggerated symptoms in one of the forms of dementia precox. When the disease has flowered, the symptoms are varied, and there are three special forms, as well as intermediate forms, which are usually easily recognized. The three relatively distinct forms are respectively the simple, or hebephrenic, the catatonic, and the paranoid.

The hebephrenic is well described as the individual who lacks emotion. He not only fails to react to situations which produce emotional states in other people but he does not exert himself in any way, and he is thought to be indolent. It is important to note that the lack of reaction is not due to a control of the affective movements but it is a real indifference to his surroundings. The events of the day pass him by without any effect. He is, therefore, described as apathetic. Even the usual bodily needs (food or drink, the necessity of emptying the lower bowel or the urinary bladder) are often neglected. Because of this the reflex mechanisms may work and the bodily functions may be carried on without the usual social inhibitions. This condition is frequently said to be an "untidy" state. There does not seem to be a great tendency to hallucination or delusion because the patient does not exhibit the motor reactions which are characteristic of these mental states.

In catatonia the principal obvious disturbances are motor, but they may differ considerably in individual patients. Two varieties with marked differences in symptoms are described. In each of the varieties, however, the symptoms may vary. In the stuporous state the patient superficially appears to be unconscious of what is happening. That he is not unconscious is shown by the fact that subsequently he may be able to recount events, even those which were relatively unimportant. The unconsiousness is in reality only a failure to react to stimuli. Not only are hunger, thirst, cold, and heat disregarded, but painful stimuli are not reacted to. There is also a failure to speak and for long periods of time the patient may be mute. These inactivities are what is called passive negativism. On the other hand, but physiologically closely associated with the passive condition, there may be other motor abnormalities. He may assume most peculiar bodily or facial attitudes. During his waking periods he may remain huddled up in a squatting position, his head between his arms and his knees. The facial muscles may be contracted or the lips twisted into grimaces. The eyelids may remain wide open, or be tightly closed. He may put his fingers into the auditory meati as if to shut out all sounds. Or he may stand rigidly in one position with arms or legs awkwardly placed. For example, he may stand in a corner of the room on one foot, with the other leg flexed so that his foot is placed below the buttocks. Allied to these postures, and at times alternating with them, there are others which are known as flexibilitas cerea. This is a state in which the patient is like wax and can be molded into whatever position is desired. The arms and legs can be arranged by another person, and the patient retains the positions as if he were a statue. Sometimes the imposed position starts the patient to carry out a series of actions which continues until the stimuli cease. An instance of this is related to eating. The catatonic may pay no attention to a call to the table, but if he is led thereto, made to sit down, he may continue to eat until the food has been consumed. Or, the handle of a floor polishing brush may be put in his hands and (especially if the nurse starts the proper movement) he will continue to polish the floor until he has gone over the whole room. In some instances he will push the polisher to and fro in the same place until his position is changed. These phenomena have been thought to be due to increased suggestibility. Allied to them and obviously due to the patient's acceptance of stimuli as suggestions to action are two states, echopraxia and echolalia, which are also found in this class of patients. Echopraxia is the senseless repetition of actions which the patient sees carried out by some one near him, and echolalia is the equally senseless repetition of sounds, mainly words and phrases, which he hears.

In catatonic excitement the picture is reversed. Instead of the placidity there is great activity. The patient is characterized by wild unreasonable actions. He may attack persons, break furniture, and there may be homicidal or suicidal attempts. At times there are monotonous movements somewhat like those in the stuporous states. These may be in talking or in gestures. He may have series of special movements, such as walking in one place a set distance, or walking backwards. Or, there may be a repetition of an arm movement, or the patient may repeat over and over one or more senseless phrases. None of these phenomena is voluntarily explained or complained of by the patient, and when questioned about them he can give no reason for them.

The third form of dementia precox is called paranoid. This state is characterized principally by the presence of delusions, usually accompanied by hallucinations. The delusions may be systematized, that is, worked over into a system which is relatively fixed, although they may be bizarre. They may, however, be unsystematized, having relatively little coherence with the other things of which the patient complains or with the events which happen to him. In some cases the delusions develop slowly and there is a gradual accumulation of the patient's ideas into a mass centering around his hallucinations and the delusion that is being formed. In other instances the delusion is quickly formed. The delusion may also change from week to week. In such cases usually it is not systematized, although often it may be definitely related to the patient's personality. The character of the delusion is obviously associated with the patient's training. Seldom, for example, do we find the delusion of being a king exhibited by an American patient. At the same time the patient may take into his delusional system items from the daily press, such as the theory of relativity. Hypnotism was more a favorite delusional element several years ago, today it is "wireless." It appears as if the least understood factors in life are those which are most likely to be contained in the delusion, perhaps because of their mystery. The patient who complained that she was dead, but subsequently at meal time complained that the food was poisoned illustrates a mixture of delusions which are incompatible. Most patients show emotional reactions to their delusions in accordance with the contents. They may be depressed if they think they are being consumed by a loathsome disease or that they are to be killed. Ideas of persecution may make them sad or resentful. In the latter case the patient may refer the origin of his trouble to some one near him and assault that person, even to the point of killing his supposed persecutor. Those whose delusions make them feel honored, wealthy, or missionary, are usually happy, although they may resent the indifference of those about them to their supposedly important status. On the other hand, those who have lived with their delusions for many years, and especially those who are dementing rapidly, may have no affective accompaniment. Since the eventual result in all these cases is that of dementia, when the patient begins to fail to attend to other ordinary occurrences he also pays less attention to his delusions.

Closely allied to the paranoid dementia precox in the way of delusion formation is that of paranoia. Paranoia is the mental disease characterized pre-eminently by a systematized, fixed delusion in an individual who does not become demented. It may start rather later in life than does dementia precox, and the patient may live for years without showing signs of mental deterioration. This means only that the illogical reasoning of the patient regarding himself does not interfere with his other mental relations. Emotionally he reacts in accordance with his ideas. His judgment regarding those matters associated with his delusional system may be poor when judged by another person, although it may be logical if the truth of the delusional idea be assumed. In matters not closely allied to his special false belief his judgment may be excellent. Sensations are rarely obtunded, but there may be hallucinations. On the other hand, as the development of the delusional system goes on, it tends to occupy more of the attention of the patient, and he may exhibit no interest in the events of the day even though they be very close to him. This form of mental disease was often called monomania, and there were as many kinds of monomania as the delusional ideas which were expressed. It was possible, however, to form groups of them, as did the English psychiatrists, in accordance with the general idea, whether religious, legal, regal, inventive, amorous, etc. In paranoia the development of the delusion appears in most cases to have a typical course. The personality basis is that of egoism. Added to the tendency to refer everything to himself, or to look at every occurrence in relation to his own happiness and welfare, there is a suspiciousness. Daydreaming may be present, and the patient may be satisfied with seclusion. The egotism, daydreaming, and seclusiveness do not enable the individual to make his way successfully in an occupation or in school, and because he is fundamentally well satisfied with himself, he refers his difficulties to his surroundings, both people and things. When, however, the surroundings are changed, the same difficulties may reappear, sometimes almost immediately and at other times only after the patient has become thoroughly accustomed to them. Along with these phenomena hallucinations occur, and these are interpreted. At first, sounds may be considered to be disagreeable, but later they may be interpreted as voices, and subsequently they may be attached to the persons with whom the patient has had difficult or uncomfortable relations. Throughout this period he may be irritable or melancholy because of the supposed activities against him. At this time he may seek revenge, or he may appeal to the courts or to other persons in authority. Or he may take personal means of chastising those who are interfering with his welfare. Rumination over his past life frequently makes him recall many things to which he has not previously paid attention as being important to him. Many items which had been separated by intervals of space or time are brought together into his systematization. He has memory falsifications, and not infrequently he picks out events which have not happened to him to fill out systematically his ideas. Later the reasons for his persecution are sought, and many times they are recognized to be due to his own importance. The delusion then becomes one of a grandiose character. Although the patient may remain irritable, he most frequently is happy, because he knows that all will come out well in the end. He knows that he will come into his rightful inheritance and honors. This last stage is said by some psychiatrists to be evidence of mental decadence or dementia, but others look at it solely as the logical development of the delusional systematization.

It will be noticed that paranoia has close resemblances to the paranoid form of dementia precox. It also has resemblances to what are called paranoid trends in many other cases. Attention may be called to those which were alluded to in the consideration of senile persons and in cases of general paralysis of the insane. Other similar states are found in middle life and in the pre-senile period, especially if the person has failed to succeed and is relatively egoistic and introspective. He refers his failure not to himself but to his lack of opportunity or to persons. He may assume an air of superiority—intellectual or social—and be recognized as "peculiar." The mental mechanism in these cases is that of many normal individuals, in hiding or disguising their failures and in excusing themselves under all circumstances. We should also recognize that the eternally suspicious, the consistent malcontents, the self-satisfied, the righteous and the reformers, the litigious, and many others have the paranoid constitution, although they may not be sufficiently unsocial or antisocial to warrant their inclusion among the insane.

Of quite a different type are those individuals who are constitutionally sad or depressed, or who are happy, or who have alternations of happiness and mental depression. These people, when they are overcome by the course of events, develop symptoms commonly called melancholia or mania. A study of the cases which superficially appear entirely different because of the dissimilarity of symptoms has shown that they have one mental disease. This has been called manic-depressive psychosis or cyclothymia, of which there are several forms.

Melancholia, or the depressed form of the disease, is characterized by a deep sadness, a feeling of unworthiness, fears, and usually a slowness of movement. The affective state is constant, whereas the reactions of the patient may vary from an almost complete cessation, which differs materially from the negativism in dementia precox, to one in which the movements are only slightly less than in the normal. Speech reactions, both vocal and writing, are slow. Vocalization may be like a whisper, and answers to questions may not begin for a minute or two. Sometimes the patient repeats over and over some stock phrases which are difficult to understand because of the low voice, such as, "What will happen to me?"; "What have I done?"; "When are they going to take me?"; "This is the end," etc. These are reactions to his depressed ideas and at times he seems to border on a delusional state, thinking, for example, that people are coming to take him away and kill him in a frightful manner because of his supposed misdeeds. He does not take in much of what is going on about him, and at times there are illusory conditions. He may describe this state as one of strangeness, or he may complain that everything is twisted or that the world is on fire. He refuses to exert himself mentally, and all such operations as calculations are done slowly and often inaccurately. The sadness and the failure to take in occurrences make him appear lacking in thought, having a paucity of ideas. There can be no doubt, however, that he is preoccupied with his ideas of sin and unworthiness, and constant attention to this would preclude thinking of other (to him, less important) matters.

Apparently opposite to the latter condition is that of mania in which state the patient pays attention to everything which stimulates his sense-organs. He reacts to them, both by general behavior and by speech. Attention is constantly shifting, and because of the concomitant speech by which he gives evidence of his ideas, his ideas appear to be numerous, rapid, and varied. If he is started on a train of thought and is not stimulated and distracted by outside noises or other sensory stimuli, he may recount, in phrases or in single words, series of ideas, or mental associations, of a varied character. At times the connections are not easy to determine, but in many cases rhyming words are given in series. In other cases there are superficial associations. This is called "flight of ideas." The emotional state is one of happiness, even though, for example, he may be slightly clad on a very cold day or have what in a normal person would be considered a very painful infection. He not only feels happy, but he thinks himself capable of the greatest things. His movements appear to be more rapid, but there is good evidence that both ideas and movements are much slower than those of a normal person, and at times nearly as slow as in depressed states. The apparent rapidity is due to their irregularity and their changes in direction.

The course of the disease and the transitional forms have indicated that melancholia and mania, as described above, are not two diseased conditions, but different manifestations of one disease. In a few cases of mania, activity, especially of speech, is noticeably lacking. In other cases of melancholia the retardation and inertia are replaced by activity. When the patient has had several attacks of the disease, all of the same type but separated by normal mental periods, the adjective "recurrent" is used to describe the condition. Thus we speak of recurrent depression or melancholia and recurrent excitement or mania. In many persons there is an alternation of the forms. At times mania may supersede melancholia overnight; at other times there may be normal periods between the two forms; and in other cases a melancholia may be immediately followed by a mania, and this in turn by a normal period. In this disease recovery is the rule in relation to the individual attack, but the disease is recurrent, the normal periods usually becoming shorter and the periods of disturbance longer as the patient grows older. The attacks may be of short or long duration. If the patient is first affected late in life the condition is usually more grave (as in involutional melancholia, which will not be discussed).

Tradition in medical practice has set apart a number of abnormal personalities from those who are feeble-minded or who have mental disease, although the personalities are mental deviates. These include the cases of hysteria, and other forms of psychoneuroses as well as those individuals who are called psychopathic personalities and constitutionally inferiors.

The symptoms which the hysteric person may exhibit are varied in the individual and diverse in groups of patients. It is not to be wondered at that there is little unanimity of opinion regarding the origin and nature of the mental state. We find one patient whose main symptom is that of inability to see or to hear, another who has convulsions, a third who has paralysis, a fourth who has loss of memory, and a fifth may be a combination of all these symptoms. At the same time, the methods of treatment are so different and so differently applied, with obviously the same results, that it is possible there may be many factors involved in the production of the hysterical symptoms. Babinski (23), for example, thinks that the hysteric is primarily a suggestible person, who may take suggestions from others as well as from himself. He says the symptoms are produced by suggestion and they can be cured by suggestion and persuasion. Rosanoff (27), while believing that hysteria is primarily a personality exaggeration, refers it to the social attitude which he calls "imposed ethics." He finds in the hysteric the tendency to make situations valuable to himself, without the feelings, as he puts them, that, "It is better to be right than to be president" (pure or aesthetic ethics), and "Honesty is the best policy" (prudence ethics). Freud believes the symptoms are due to conditions in the unconscious, referring them to sexual situations which have occurred prior to puberty. Whichever explanation or personality trait is thought to be most inclusive, there can be little doubt that the symptoms are of the nature of defensive mechanisms. The hysterical person seeks comfort, approbation, release from difficult or dangerous situations. In seeking these things, he, consciously or unconsciously, adopts modes of reaction which will bring them about. What he exhibits as symptoms depends upon his knowledge of conditions. He adopts suggestions, as we have said above, from within and without, which add to his disabilities but which bring him the relief. He may, therefore, become paralyzed if he has to march with his regiment to the front, or he may have tremors so that he cannot handle his rifle, or he may become deaf so that he cannot understand commands. In the civilian status the symptoms also depend upon surroundings. Convulsions, tics, anaesthesias, speech disturbances, may be shown, but all are of characters which are recognizable as not being dependent upon organic disease. There may be peculiar memory disturbances, part of the past being blotted out—either anterograde or retrograde amnesias. The patient may "forget" his past and there may be multiple personalities.

The relationship of hysteria and malingering has often been considered. Babinski believes malingering to be essentially different from hysteria, and this should also be the conclusion of the psychoanalytic school, which places emphasis upon sexual accidents in childhood as the cause of hysteria. However, the symptoms in both conditions may be the same. The difference is largely one of obvious motive. The hysterical motive is not on the surface; the motive of the malingerer is easy to see. In some cases it is apparent that the hysteric seeks examination and appears anxious to overcome his abnormalities. The malingerer, on the other hand, seldom desires to recover until the occasion for his disability is passed. Nevertheless the differential conditions are not constant, and there appears no valid reason for separating these two conditions.

To estimate the prevalence of these conditions is almost impossible. Given certain difficult or dangerous circumstances, it is apparent that symptoms may develop which may be considered those of malingering or of hysteria. During the war there were many cases in which fear brought about the classical symptoms, and in civil life it has been estimated that more than 50 per cent of physical complaints are imagined. Wilson, citing the data collected in connection with English railroad accidents, shows that about 70 per cent of those who had no observable physical injury following these accidents claimed damages from the railroad companies for various supposed injuries, including mental shock. Most of the war cases, as well as most of the railroad cases, recover as soon as the occasion for the symptoms has passed. Following the armistice there was an immediate recovery of many army cases of hysteria; following a favorable jury verdict many of the railroad cases have a return to a normal state.

Among other psychoneurotics are the neurasthenics and the psychasthenics. These make up the large group of patients who constantly seek aid from physicians, the church, quacks, patent medicines, and many other people and things. They recite symptoms to any one willing to listen, go to the cure suggested, acquire new symptoms from others who recite their own ills. They are absorbed and always anxious about their conditions; they often manage to carry on business and social activities but take their pleasures mostly in observing and recounting their various ills. Among this class are some who become greatly depressed, inattentive, and inactive, and they resemble, or they may actually be, simple depressions of the nature of manic-depressive psychosis. Obsessions, such as kleptomania and pyromania, are found in some patients and there are fears of different kinds—of closed, open, or high places. There are many different kinds of obsessions, and there are many items of which these abnormal individuals have fears. Stage fright, difficulties in making certain of decisions, and many states of doubt are mild conditions of psychasthenia which may subsequently develop into the more intense phobias, obsessions, and anxieties.

Throughout this chapter there have been recounted the underlying kinds of personalities which develop into the better known mental abnormalities. Quite frequently, however, we see the paranoid personality, with suspiciousness and egotism, remain at an acceptable social level. We also see the cyclothymic personality, with depression, exaltation, inactivity, or talkativeness, go no farther than the state which others call "moody." Impulsive persons, corresponding to the more marked catatonics, may become antisocial on occasions, but they may react only by being uncertain towards their employers and to others with whom they must live. It is likely that the impulsive criminal is of the latter class. Other criminals have doubtless more of the nature of the underlying dementia-precox personality.

It should also be recalled that the personality traits of most individuals change because of recognizable physical diseases. The cheerfulness of the pulmonary tuberculous, while not constant, is often apparent. The mental distress, apart from pain, of the person with disease of the heart or of gastro-intestinal system is equally obvious. So also, because of general or localized disturbances of the brain, as in encephalitis lethargica, there may be personality variations. Unfortunately we cannot associate the mental abnormalities with definite cerebral changes.

REFERENCES

GENERAL

1. Beers, C. W. A mind that found itself. New York: Longmans, Green, 1908. Pp. ix+363.
2. Church, A., & Peterson, F. Nervous and mental diseases. Philadelphia: Saunders, 1911. Pp. 932.
3. Conklin, E. S. Principles of abnormal psychology. New York: Holt, 1927. Pp. vii+457.
4. Jelliffe, S. E., & White, W. A. Modern treatment of nervous and mental diseases. (2 vols.) Philadelphia: Lea & Febiger, 1913. Pp. xvi+867; viii+816.
5. Jung, C. G. Psychological types. New York: Harcourt, Brace, 1926. Pp. xxii+654.
6. McDougall, W. Outline of abnormal psychology. New York: Scribner's, 1926. Pp. xiii+566.
7. Morgan, J. J. B. The psychology of abnormal people. New York: Longmans, Green, 1928. Pp. vii+627.
8. Pressey, S. L., & Pressey, L. C. Mental abnormality and deficiency. New York: Macmillan, 1926. Pp. xii+356.
9. Rosanoff, A. J. Manual of psychiatry. New York: Wiley, 1927. Pp. xvi+697.
10. Störring, G. Mental pathology in its relation to normal psychology. London: Sonnenschein; New York: Macmillan, 1927. Pp. x+298.
11. Taylor, W. S. Readings in abnormal psychology and mental hygiene. New York: Appleton, 1926. Pp. 822.
12. Wallin, J. E. W. Clinical and abnormal psychology. Boston: Houghton Mifflin, 1927. Pp. xxii+649.
13. Wells, F. L. Mental adjustments. New York: Appleton, 1922. Pp. xiii+331.
14. White, W. A. Foundations of psychiatry. Washington, D. C.: Nerv. & Ment. Dis. Publ. Co., 1921. Pp. ix+136.
15. ———. Mechanisms of character formation. New York: Macmillan, 1916. Pp. iv+342.

FEEBLE-MINDEDNESS

See Nos. 2, 3, 8, 9, 12.

16. Barr, M. W., & Maloney, E. F. Types of mental defectives. Philadelphia: Blakiston, 1920. Pp. ix+179.
17. Goddard, H. H. Feeble-mindedness, its causes and consequences. New York: Macmillan, 1914. Pp. xii+599.
18. Seguin, E. Idiocy and its treatment by the physiological method. New York: Teach. Coll., 1907. Pp. 202.
19. Tredgold, A. F. Mental deficiency. London: Baillière, 1908. Pp. xviii+391.

MAJOR PSYCHOSES

See Nos. 1, 2, 4, 9, 11, 14

20. Bleuler, E. Textbook of psychiatry. (Trans. by A. A. Brill.) New York: Macmillan, 1924. Pp. xviii+635.
21. Cotton, H. A. The etiology and treatment of the so-called functional psychoses. *Amer. J. Psychiat.*, 1922, 2, 157-210.

22. **Kraepelin**, E. Textbook of psychiatry. Vol. 3, Pt. 2. Dementia praecox and paraphrenia. (Trans. by R. M. Barclay.) Edinburgh: Livingstone, 1926. Pp. xii+328.

PSYCHONEUROSES, ETC.

See Nos. 3, 4, 9, 13, 14.

23. **Babinski**, J. My conception of hysteria and hypnotism. *Alienist & Neurologist,* 1908, **19**, 1-30.

24. **Janet**, P. Major symptoms of hysteria. (Trans. by C. R. Corson.) New York: Macmillan, 1907. Pp. 345.

25. **MacCurdy**, J. T. Problems in dynamic psychology. New York: Macmillan, 1922. Pp. xv+383.

26. ————. War neuroses. Cambridge: Cambridge Univ. Press, 1918. Pp. ix+132.

27. **Rosanoff**, A. J. A study of hysteria. *Arch. Neur. & Psychiat.,* 1919, **2**, 419.

28. **Yealland**, L. R. Hysterical disorders of warfare. London: Macmillan, 1918. Pp. 259.

MENTAL TYPES

See Nos. 5, 6, 13, 15, 20, 24.

29. **Brill**, A. A. Fundamental conceptions of psycho-analysis. New York: Harcourt, Brace, 1921. Pp. vii+344.

30. **Ellis**, H. The criminal. New York: Scribner's, 1890. Pp. viii+337.

31. **Galton**, F. Hereditary genius. New York: Appleton, 1891. Pp. xii+390.

32. ————. Inquiries into human faculty. London: Macmillan, 1883. Pp. xii+387.

33. **Gross**, H. Criminal psychology. Boston: Little, Brown, 1918. Pp. xx+514.

34. **Healy**, W. Mental conflicts and misconduct. Boston: Little, Brown, 1917. Pp. xi+330.

35. **Hoch**, A. Constitutional factors in the dementia praecox group. *Rev. Neur. & Psychiat.,* 1910, **8**, 463-474.

36. **Meyer**, A. The role of the mental factors in psychiatry. *Amer. J. Insan.,* 1908, **65**, 39-56.

37. **Rivers**, W. H. R. Conflict and dream. New York: Harcourt, Brace, 1923. Pp. xi+194.

38. **Sullivan**, W. C. Crime and insanity. London: Arnold, 1924. Pp. vii+259

CHAPTER 22

GENERAL STATISTICAL PRINCIPLES

TRUMAN L. KELLEY
Stanford University

AND

EUGENE SHEN
China Institute in America

Statistical method is an instrument for summarizing and analyzing data obtaining from a series of observations. The statistical treatment of such data serves a twofold purpose: it yields a descriptive account of observed facts and at the same time affords a basis for an estimate of some unobserved events. Inasmuch as a description of the observed is accurate, the method is always valid. But inasmuch as the unobserved cannot be predicted with certainty, it is frequently fallible. When a particular description of observed facts is generalized for predictions of unobserved events, its status becomes that of a hypothesis. A hypothesis is true or false according as predictions based upon it hold or fail. If a prediction fails, it will be found to have been based upon an imperfect description which left certain relevant factors out of account. The evaluation of relevancy of the various factors in a situation is not infrequently a matter of prestatistical analysis upon which is determined the statistical procedure.

The use of statistical treatment in research is not altogether incomparable to the application of simple arithmetic to commonplace problems, such as finding the cost of several pencils from the cost of each. The successful solution of this problem requires three essential conditions: the number of the pencils and the cost of each must be accurately ascertained, the relation between the different items of information must be properly apprehended, and the computation must be correctly carried out. The place of statistical technique in research is analogous to the part played by computation in our hypothetical problem. An irrelevant treatment of data need never be marred by any error as far as the strictly statistical technique is concerned. Neither does accurate arithmetic redeem the sins of careless observation. On the other hand, however, a suggestive insight into a problem will lack exact formulation and proper verification if the requisite statistical methods are not at command. In other words, familiarity with statistical technique is indispensable, though by no means sufficient, to investigators in certain branches of science.

The specific function of statistical treatment is to group and analyze a set of data so that certain constant features are made apparent. The actual calculation of the statistical constants is a strictly mechanical though sometimes highly complicated procedure. But the method of treatment should be selected with a view to suit the problem. An ill-chosen treatment is vicious not so much because of its futility toward the solution of a given problem as because of its likelihood to lead to misinterpretation of the results. While there is always opportunity for the exercise of ingenuity to devise new techniques in attacking new problems, certain concepts and methods of procedure are fundamental as well as generally applicable, and have been extensively employed in psychological investigations.

Statistical Series

A set of statistical data is a series of observations in respect to some common characteristic. A series can be classified as one of the following four types, and there are certain statistical concepts and techniques which are peculiar to each.

1) *Geographical Series.* Geographical series are the two-dimensional or most important subdivision of the more inclusive type which can be described as "spatial." They are peculiarly amenable to graphic treatment, and the map is especially useful. Statistics of census returns and of climatic conditions are examples of geographical series.

2) *Temporal Series.* The concepts of increment and ratio are peculiar to temporal series. It is generally a characteristic of the temporal series that the relative value of a function is much more significant than the absolute value. Accordingly, in treating temporal series, we frequently make use of index numbers, periodograms, and growth curves, and employ geometric means instead of arithmetic means. Examples of temporal series are data showing trends of prices, diurnal changes in temperature, physical and mental growth, learning, forgetting, and fatigue.

3) *Qualitative Series.* In qualitative series, we have frequencies in discrete categories, and the basic concept in statistical treatment is the variability of a frequency. They lead to the consideration of the probability of frequencies and of proportions. If the qualities are ordered in a linear manner, such as students' grades in terms of A, B, C, and D, the series may be converted into a quantitative series by assigning a numerical value to each of the letter grades. The equivalence of successive units in such converted values cannot be established except by means of assumptions in regard to the form of distribution.

4) *Quantitative Series.* Quantitative series lead to distributions of frequencies along numerical scales of certain measures. They are the most fertile for numerical treatment and their consideration will comprise the larger part of the following pages. Such statistical constants as give measures of central tendency and of dispersion relate to quantitative series.

Errors in Statistical Data

In applying statistical treatment to a set of data, it is important to note the source of the material in respect to the method of observation and the error to which it is liable. Errors in statistical data are either errors of observation or errors of sampling, and in both cases they may be chance errors or systematic errors.

Observation errors of a chance order are present in any set of data. They occur as long as the instrument of measurement is short of perfection. They are errors which in the long run tend to compensate each other when yielding measures of central tendency. For this reason they are sometimes called "unbiased" errors, but the lack of bias does not extend to measures of dispersion and correlation. In simple physical measures such as height and weight, observation errors are usually so small that for all ordinary purposes they are negligible. In the field of psychology, such direct measurements as that of reaction-time come under the same category. Most psychological measurements, however, are measurements of quantities which are less stable. The scores on a mental test, for instance, are liable to a great variety of errors which are by no means negligible. A second test will show considerable fluctuations in the scores of the different individuals, although the average for the group may remain relatively constant. When a series of observations is repeated, the size of such chance errors can be calculated, and is usually given in terms of the reliability coefficient, which will be discussed later.

Systematic errors of observation are not unavoidable if the investigator exercises sufficient care and judgment. They are biased errors due to some constant factor and lead to distorted results. Examples of instruments and methods entailing systematic errors are meter sticks which are shorter than they are supposed to be, chronoscopes which record a longer time than the actual interval, questionnaires which suggest certain answers rather than their opposites, and psychophysical experiments in which the darker of two grays consistently receives more illumination.

Sampling errors are involved when observation is made on a relatively small number of cases which are taken as representatives or samples of a larger group. Chance error of sampling will be discussed at greater length later in the chapter. A numerical estimate of their magnitude can be derived from the number of cases observed and the variation among them. Sample errors of a systematic order, however, will not show itself automatically in the course of statistical treatment. They are due to selective, as opposed to random, sampling. If, for instance, one is studying the mental ability of 12-year-olds in general and limits the observation to the fourth grade, or if one studies Nordics in the U. S. Army draft and thinks he is studying Nordics in general, a systematic error of sampling is obviously present. The magnitude of the error is proportional to the limitations of the selected group.

It will be interesting to note that the same error may be considered an error of observation in one connection and as an error of sampling in another connection. If we consider the score of an individual on a mental test as a measure of a constant ability, there is an error of observation attributable to the imperfections of the test and inadequate control of the conditions. If, on the other hand, we regard the test score as a measure of a response which is a sample of responses to comparable situations, then by far the larger portion of the error is attributable to sampling.

Tables and Graphs

Tabulation of data is usually the first step in statistical treatment. While the choice of the unit is made prior to any statistical treatment, tabulation often necessitates grouping where the length of the interval has to be decided. The general practice is to divide the total range into from ten to twenty equal intervals. Much, of course, depends upon the size of the population as well as the range it covers, but the more significant determining factor is the problem. If, for instance, one is interested in teachers' practice or ability of discrimination in grading pupils upon a percentage basis, where multiples of five are expected to be in special favor, it would obviously be absurd to group the data in intervals of five points, a perfectly reasonable procedure under other circumstances.

It is unfortunate that the practice of labeling the class intervals has not been uniform and frequently leads to confusion. The label sometimes denotes the mid-point of the interval, sometimes the lower limit, and sometimes the upper limit. Thus one is not always sure whether a class interval labeled age 12 covers a range between ages 11.0 and 12.0, between 11.5 and 12.5, or between 12.0 and 13.0. Careful investigators always make this point clear, not only to avoid ambiguity to the reader but also to prevent errors in subsequent treatment.

Wide option usually exists in the details of presenting even very simple data. In general two types of tables are distinct. A table of more or less raw data without abridgment serves the purpose of general reference, while tables for special purposes are so arranged as best to emphasize certain points of the study. Good tables frequently accomplish a presentation for multiple purposes in a single arrangement.

Tables often serve as bases for plotting graphs. Our response to pictorial representations being more direct and depending less upon special acquired associations than our response to verbal symbols, graphs give us clearer and stronger impressions than mere figures in tables. They are employed not only for presenting data to the reader but also for inspection by the investigator himself preliminary to more elaborate analysis.

Provided a graph is accurately plotted, it always presents a true picture of the facts. A graph, for instance, showing the amount of work done during successive units of time is considered by itself just as true as one showing the amount of time taken by successive units of work, though there may be a question as to which is more advantageous for a given purpose. A conversion of the unit in terms of which the data are plotted often results in a very radical alteration of the appearance of the picture.

A very common graph is one which represents the distribution of frequencies. In such graphs, the amounts in a certain score are commonly plotted as abscissas, and the frequencies as ordinates. Two forms may be distinguished. The *histogram* consists of a series of rectangles erected upon successive portions of a base line. If, instead of the rectangles, the successive heights representing the frequencies are plotted and the points joined by straight lines, it becomes a *frequency polygon.* The frequency polygon suggests the idea of continuity whereas the histogram plainly represents proportions in the various classes. A modified form of the histogram is the *bar diagram* which, with spaces between the classes, represents discrete elements.

If the coefficients of the successive terms of a binomial expansion are plotted as the frequencies of a distribution, the resulting graph will present characteristics common to many frequency distributions. By increasing the exponent of the binomial expanded, the distribution as graphically represented attains more smoothness. Such smoothing does not reach a limit until the changes from one frequency to the next are infinite in number and infinitesimal in amount. The resulting curve appears continuous and is called the curve of the *normal probability distribution.* We shall return to the characteristics of the normal distribution curve as we discuss the various statistical constants.

Instead of plotting the separate frequencies in successive intervals and obtaining a curve of frequency distribution, one may plot the cumulative frequencies up to successive limits and obtain what is called an *ogive curve.* Since the frequencies cumulate at successive limits, the ogive curve always has a positive slope. Its appearance is thus quite distinct from that of a frequency distribution. The curve of the *phi gamma function* in psychophysics is an ogive curve converted from a normal distribution.

Somewhat similar in appearance to the ogive curve are curves representing the trend of increase of a given measure in the course of time. Units of time are plotted as abscissas, while the ordinates represent such quantities as of height, weight, and measures of intelligence or achievement. Curves of this type which are frequently met in the study of psychology are either *growth curves* or *learning curves.* It will be seen that by plotting the increments instead of the gross measures, the curve may revert to the general appearance of the frequency distribution.

In plotting graphs, data are sometimes smoothed by the method of *moving averages,* i.e., by averaging the ordinate at each abscissa with those of the neighboring abscissas. Such smoothing is employed to help show general trends which may otherwise be obscured by minor fluctuations. But it must not be forgotten that minor fluctuations are sometimes themselves significant and should not be overlooked. It is hardly necessary to add that (with the exception of the mode) statistical constants are never to be determined from smoothed data.

At times instead of joining points with straight lines, a continuous curve is drawn through them. This is done to correct the limitations of observation which often forces continuous phenomena into discrete units. Reasonable exactness in this procedure, however, would require the derivation of an equation to fit the data. The choice in type of equation is arbitrary and there is a wide range to choose from.

A frequency distribution presented either in tabular or in graphic form is usually so full of details as to make interpretation difficult. For presentation in a more concise form, we employ certain numerical derivatives of the data which are called statistical constants. The remaining part of the present chapter will be largely devoted to a general discussion of these constants. Limit of space precludes giving detailed steps in the procedure, for which the reader is referred to textbooks on statistical method.

MEASURES OF CENTRAL TENDENCY

Measures of central tendency are often spoken of as "averages," but "the average" is also used for the *arithmetic mean,* which is one particular measure of central tendency. Other measures of central tendency are the *median,* the *mode,* the *geometric mean,* and the *harmonic mean.*

The arithmetic mean, or briefly the mean, is the sum of the separate measures divided by the number of cases. It is that measure for the group from which the square of the deviation of each individual measure will add to a minimum. Graphically considered, the mean is the abscissa at which the area under the curve of distribution will exactly balance itself. It is thus sensitive to any change of a single measure.

To calculate the mean, it is necessary only to add the measures and divide the sum by the population. Thus:

$$M = \frac{\Sigma X}{N} \qquad [1]$$

where X designates any individual measure, Σ a summation, N the number of cases or the population, and M the mean. But usually the measures are grouped according to magnitude, and it will also shorten calculation and facilitate the derivation of other measures if a magnitude guessed to be near the mean is used as an arbitrary origin from which deviations of the separate measures are computed to make the necessary correction. Using f to designate the frequency within a class interval and ξ the deviation in terms of the number of intervals of a class index from the arbitrary origin or guessed mean, we have:

$$M = \text{Arbit. Orig.} + \frac{(\text{Class interval})\ \Sigma f\xi}{N} \qquad [1a]$$

The use of an arbitrary origin does not in any way affect the result of calculation, but grouping does. Errors introduced by grouping, however, are insignificant, if grouping is not too coarse and if frequencies within the groups do not consistently bunch near either the lower or the upper boundary.

The median is the value of the mid-most measure, the abscissa which divides the area under the curve of distribution into two equal halves. If deviations are taken from the median of the separate measures regardless of sign, they give a sum less than that of deviations taken from any other point.

The median of a distribution is simply obtained by taking the value of the mid-most measure if the total number of cases is odd or by averaging the two mid-most measures if the number is even. When data have been grouped, calculation of the median assumes that the frequencies within the interval in which the mid-most measure lies are evenly distributed in the form of a perfect rectangle. The median will be in error in proportion as the obtained condition differs from this assumption.

Unlike the mean, the median is insensitive to changes in the extremes of a distribution. In fact, it is insensitive to any fluctuation not crossing the boundary of the class interval in which it is located. For this reason, and also for the ease of its derivation, some investigators have often preferred the median to the mean as a measure of central tendency. It is obvious, however, that this insensitivity of the median is sometimes a weakness rather than strength. Besides, only when a distribution is very peaked in the neighborhood of the mode and has one or two long tails (such a distribution is termed leptokurtic, as we shall see in the section on curve-fitting) is the median actually more reliable than the mean. It should also be noted that the median of a combined distribution cannot be derived from the medians of the separate distributions.

Since the median is the value below which 50 per cent of the measures lie, it is the 50-*percentile*. By generalizing about the 50-percentile, it is very simple to define or derive any percentile. The percentiles are very useful measures, but we should, of course, refer only to the 50-percentile as a measure of central tendency.

The mode is the value in a series at which the greatest frequency lies. Graphically, it is the abscissa at which the curve of distribution attains a maximum height. Minor modes are occasionally found; they are places at which the height of the curve is greater than that on either side and yet less than the greatest height of the whole curve.

An approximate location of the mode can be readily made by inspecting the table or curve of frequency distribution; but an exact determination is rather difficult. If irregularities are present in the distribution, the major mode is usually obtained after smoothing the data by the moving average method. A more precise but more laborious method is to obtain the maximum of a curve mathematically fitted to the data.

Another method of determining the mode is calculation from the mean and median, utilizing the following relation which holds between them for a majority of distributions:[1]

$$\text{Mode} = \text{Mean} - 3.03\ (\text{Mean} - \text{Median}) \qquad [2]$$

The formula shows that when the mean and median coincide the mode is equal to both. This is obviously an indispensable condition for perfect symmetry of a distribution. The curve of the normal probability distribution previously mentioned is a special case of symmetrical distributions.

The presence of minor modes is an indication of three possible conditions. It may be simply due to accidental irregularities of a small population, in which case the minor modes are plainly of no significance. But they may be expressions of some true periodicity such as in the case of teachers' predilection to multiples of five in grading pupils on a percentage basis. Periodicity, however, may not be revealed and may be given the appearance of chance irregularity if the data have not been appropriately groupped. On the other hand, persistent multimodality transcending regrouping is a clear indication of heterogeneity of the population, and some character may be found by which the measures can be divided into distinct classes not continuous with each other.

---

[1] This formula is an approximation of the more accurate relation, $Mo = M - \frac{M - Mdn}{c}$, where c is closely equal to $.3309 - \frac{.0846\,(M-Mdn)^2}{\sigma^2 - 9\,(M-Mdn)^2}$. Neither formula, however, is applicable to skew L-shaped or bimodal U-shaped distributions.

Besides the mean, the median, and the mode, we may mention another measure of central tendency which is of more restricted application. If measures are ratios which express each item in terms of the preceding, such as annual price indices relative to the preceding year, the arithmetic mean would be difficult of proper interpretation, and the significant average to use is the *geometric mean,* which is defined as the *n*th root of product of the separate measures, *n* being the population. The calculation of the geometric mean is facilitated by the use of logarithms.

Measures of central tendency are significant and useful as representatives and are usually the best single measures for a group. But if the individual measures are not of equal importance nor of equal accuracy, averages calculated from the raw measures are not the best measures for the group. The best answer to a mathematical problem is not the average of the right and wrong ones, nor is the true average height of man the average of the heights of different groups by race or nationality. The correctness of the answers in the first instance and the size of the groups in the second should be taken into consideration. In other words, the measures should be weighted before averages are calculated. In combining averages of different groups, they should be weighted directly as the number of cases in each group, and, in combining measures of varying reliability, they should be weighted inversely as the squares of the errors.[2]

A special case of weighted average is the *harmonic mean,* which is the arithmetic mean of a set of measures weighted according to their reciprocals. Thus:

$$H.M. = \frac{\Sigma\left(\frac{1}{X}\right)X}{\Sigma\left(\frac{1}{X}\right)} = \frac{N}{\Sigma\left(\frac{1}{X}\right)} \qquad [3]$$

In a relay, for instance, in which the successive runners cover the same distance, they run for different lengths of time proportional to the reciprocals of their rates (distance per unit time). Consequently the average rate of the whole course is given not by the arithmetic mean but by the harmonic mean of the individual rates. The harmonic mean is usually defined as the reciprocal of the arithmetic mean of the reciprocals of the measures. A simple algebraic operation will verify that it is equivalent to the definition which we have offered. It is an invariable relation between the various measures of central tendency that the harmonic mean is less than the geometric mean, which is in turn less than the arithmetic mean.

---

[2] The exact error of any measure is of course unknown, or else it could have been corrected. But the probable size of the errors is often available. For methods of estimating the size of errors, see sections on Errors of Sampling and Reliability and Validity.

Measures of Dispersion

Distributions having the same average may differ markedly in the dispersion of the measures. Given a definite amount of wealth for a constant population, the mean wealth of the individuals in the group will be the same regardless of the way the wealth is divided among them, but the scatter may vary all the way from zero, i.e., all individuals possessing exactly the same amount, to any magnitude which the total wealth will allow. To know the exact degree of scattering, it is necessary to derive some measure of dispersion. Measures of dispersion may be based upon the deviations of different measures from the mean, or the squares of such deviations, or upon deviations from some other measure of central tendency, or upon the range between certain two measures.

The simplest measure of dispersion is the total range, the difference between the lowest and the highest score. But it is not a reliable measure, as it depends altogether too much upon the extremes of a distribution. A much better measure which has been commonly used is one based upon the interquartile range, or the difference between the 25th and 75th percentiles. The *semi-interquartile range,* or one-half of the interquartile range, is customarily called the *quartile deviation,* though strictly it is not a measure of deviation from the average. It is also sometimes incorrectly called the probable error, of which more will be said in the following section.

The interquartile range is easily interpreted as covering the central 50 per cent of the cases. But a measure of dispersion may be obtained by taking the difference between any two percentiles. It has been shown (12)* that for a normal distribution the range between the 6.917th and 93.083th percentiles has the maximum reliability. A range but slightly different from this and having nearly as great reliability is that between the 10th and 90th percentiles. Its calculation and interpretation is as simple as the interquartile range, but it requires 42 per cent less cases than the quartile deviation to secure the same reliability. The designation D is used for this very serviceable interpercentile range.

The *mean deviation,* also called the *average deviation,* is the sum of the absolute values of the deviations of the separate measures from the mean, divided by the population. It is usually calculated from the mean, but it is smaller when calculated from the median than from any other point. The term "mean deviation from the median" should be fully expressed if it is so calculated. The mean deviation is not difficult to calculate and is a rather reliable measure. But if the greatest accuracy is demanded, or if correlation is part of subsequent treatment, it is necessary to obtain the *standard deviation.*

---

*The bibliographical references for Chapters 22 and 23 are given at the end of Chapter 23.

In connection with the standard deviation, it will be profitable to make use of the concept of moments, which is basic in the derivation of a number of important measures. Designating a gross score by X, a deviation from the mean by x, the population by N, and the coefficient of a moment of the nth order by μ_n, we have the following general equation as a definition of moments:

$$\mu_n = \frac{\Sigma x^n}{N} = \frac{\Sigma (X-M)^n}{N} \qquad [4]$$

The mean can then be defined as the point around which the first moment is zero. The standard deviation, regularly designated σ, is the square root of μ_2, i.e., the square root of the mean of the squares of the deviations from the mean. The square of the standard deviation has been called the *variance,* a term which will sometimes be found convenient to use. The third and fourth moments are useful in curve-fitting and in deriving measures of *skewness,* or lack of symmetry, and *kurtosis,* or peakedness, of a distribution.

The standard deviation is by far the most significant and reliable measure of dispersion. Like the mean, it is usually more convenient to calculate by using an arbitrary origin and correcting for the difference, without in any way reducing its accuracy. Thus:

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = (\text{class interval})\sqrt{\frac{\Sigma \xi^2}{N} - \delta^2} \qquad [5]$$

where δ designates the difference between the mean and the arbitrary origin, the other symbols having previously been defined. The effect of grouping, however, is not as negligible in calculating the standard deviation as in the case of the mean. For calculation after grouping employs the mid-point of an interval for each measure within it, whereas the actual distribution of frequencies in an interval generally yields more than 50 per cent of the measures between its mid-point and the mean of the entire distribution. This difference entails a systematic overestimation of the total of the squared deviations from the mean. In case grouping is coarse, the error may be substantial and should be corrected. For distributions which approach the base line asymptotically at both extremities, the square of the standard deviation calculated after grouping has been found to be too large by 1/12 of an interval. This correction is known as Sheppard's correction.

In a normal distribution the various measures of dispersion are related in the following way:

$$\left.\begin{array}{rlrlrl} Q = & .84535 & M.D. = & .67449 & \sigma = & .26315\,D \\ M.D. = & 1.1829 & Q = & .79788 & \sigma = & .31129\,D \\ \sigma = & 1.4826 & Q = & 1.2533 & M.D. = & .39015\,D \\ D = & 3.8001 & Q = & 3.2124 & M.D. = & 2.5631\ \sigma \end{array}\right\} \quad [6]$$

ERRORS OF SAMPLING

The measures of dispersion are primarily descriptive measures of observed data. But they also fulfil another function. If we are confronted with the task of estimating the values of certain individual measures of which nothing is known except that they belong to a given distribution, we shall naturally use some measure of central tendency of that distribution for the best estimate. We may use the mode because it has less chance of being in error than any other value, the median because its chance of being an overestimate is equal to its chance of being an underestimate, or the mean because in the long run the squares of the amounts by which it is in error will be less than those of any other value. When measures of central tendency are used as estimates for individual measures, the errors of estimate are given by the measures of deviation.

Any obtained distribution is usually but a *sample* of a large, sometimes unlimited, population of measures of the same category to be found in nature. (The term "population" as just used refers to the large number of unobserved as well as observed cases, and not to N, which is the "population of the sample.") It is frequently of theoretical interest as well as practical importance, in using the statistical constants of the sample as estimated values for the population, to have an idea of the extent to which the estimates are liable to be in error. In other words, we desire to have some measure of the *sampling error* of the obtained constants. The most generally available and most reliable measure of sampling error is the *standard error,* which is identical with the standard deviation except that it has a predictive instead of a descriptive significance. The same symbol σ is used in both cases.

To derive the standard error of an obtained constant, say, of the mean of a given distribution, then, is to estimate the standard deviation of an imaginary distribution of a set of similarly derived means taken as gross measures. To make such an estimate, it is necessary to take the deviations obtained from the given distribution as approximations to deviations of the unavailable measures from the imaginary mean of means, and to retain only the squares of such deviations while discarding as negligible all their cross products which would add close to zero according to the product theorem.[3] Estimates obtained by such approximations of course cannot be infallible, but they give a satisfactorily discriminative indication of the reliability of the obtained constants.

As the mean and the standard deviation are based upon moments, their standard errors are likewise derivable from the standard error of moments. The standard error of any moment is given by

---

[3] The product theorem states that the sum of products of measures which are independent of each other and whose sum or average is zero is equal to zero. See section on Correlation.

$$\sigma_{\mu_n} = \sqrt{\frac{\mu_{2n} - \mu_n^2}{N}} \qquad [7]$$

From this we immediately obtain the standard error of the mean, the origin about which the first moment is zero, and of the variance or the square of the standard deviation, which is the second moment:

$$\sigma_M = \sigma_{\mu_1} = \sqrt{\frac{\mu_2 - \mu_1^2}{N}} = \sqrt{\frac{\mu_2}{N}} = \sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}} \qquad [8]$$

$$\sigma_{\sigma^2} = \sigma_{\mu_2} = \sqrt{\frac{\mu^4 - \mu_2^2}{N}} = \sqrt{\frac{\mu_4 - \sigma^4}{N}} \qquad [9]$$

The standard error of the standard deviation is derived from σ_{σ^2}. It is

$$\sigma_\sigma = \frac{1}{2\sigma}\sqrt{\frac{\mu_4 - \sigma^4}{N}} \qquad [10]$$

In a normal distribution, which is *mesokurtic,* i.e., $\mu_4 = 3\mu_2^2$, we have

$$\sigma_\sigma = \frac{1}{2\sigma}\sqrt{\frac{2\sigma^4}{N}} = \frac{\sigma}{\sqrt{2N}} \qquad [11]$$

As moments higher than the second are ordinarily not calculated, the last formula is generally used for all distributions. When kurtosis is extreme, however, the error thus introduced will not be negligible.

The standard error of the median, of any other percentile, or of any interpercentile range, has as its basis the standard error of F, the frequency in a given category,

$$\sigma_F = \sqrt{Npq} \qquad [12]$$

and of p, a proportion,

$$\sigma_p = \sqrt{\frac{pq}{N}} \qquad [13]$$

where N is the population, and p and q are proportions, respectively, of frequencies lying within and without the given category, thus $p + q = 1$. From these formulae, the standard error of any percentile is readily derived. Designating the range of the class interval in which the percentile lies by i, the frequency within that interval by f, and the proportion of frequencies below and above the percentile by p and q, respectively, we have for the standard error of P_p, the 100p-percentile, or the percentile having p proportion of cases lying below it:

$$\sigma_{P_p} = \frac{i\sqrt{Npq}}{f} \qquad [14]$$

Since the median is the 50-percentile,

$$\sigma_{Mdn} = \sigma_{P_{.50}} = \frac{i\sqrt{N\frac{1}{2}\frac{1}{2}}}{f} = \frac{i\sqrt{N}}{2f} \qquad [15]$$

The standard error of any interpercentile range may also be derived. From it the standard errors of the quartile deviation and the 10-90 percentile range in a normal distribution are calculated to be as follows:

$$\sigma_Q = .7867\frac{\sigma}{\sqrt{N}} = 1.166\frac{Q}{\sqrt{N}} \qquad [16]$$

$$\sigma_D = 2.279\frac{\sigma}{\sqrt{N}} = .889\frac{D}{\sqrt{N}} \qquad [17]$$

The 10-90 percentile range thus has a considerably less percentage error than the quartile deviation. But the standard deviation is more reliable than either, since

$$\sigma_\sigma = \frac{\sigma}{\sqrt{2N}} = .707\frac{\sigma}{\sqrt{N}} \qquad [18]$$

It will have been observed that the standard error of a constant is generally in inverse proportion to the square root of the population of the sample from which the value is obtained. In other words, the larger the sample, the more representative the distribution, and the more reliable the constants.

To interpret the standard error in terms of probability, one assumes a normal distribution of the errors. The chances for a difference between the obtained constant and the true value to exceed any amount are then equal to the frequencies lying beyond that amount of deviation, as given in tables of the normal probability integral. Thus the probability of a deviation exceeding one standard error is .3174, of it exceeding twice the standard error .0456, and of it exceeding three times the standard error .0027.

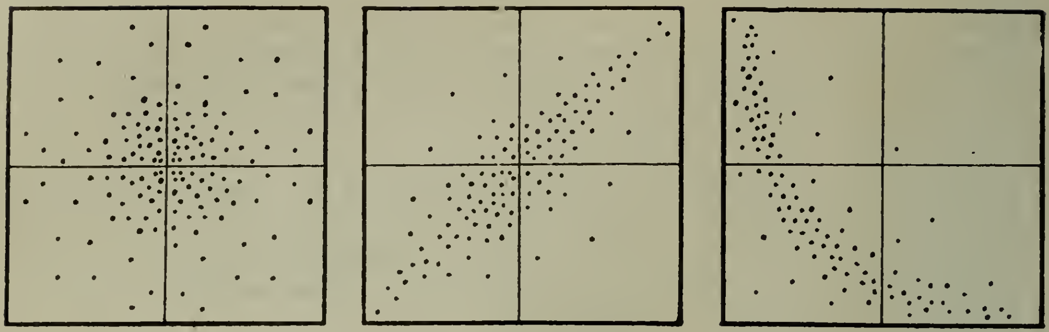
It will be recalled that in a normal distribution one quartile deviation, which is equal to .67449 of a standard deviation, above and below the mean, covers exactly 50 per cent of the cases. A measure of sampling error comparable to the quartile deviation is the *probable error*. An obtained measure thus has an even chance of lying within, as without, one probable error either way from the true measure. Because of this apparent simplicity of interpretation, writers have often preferred the probable error to the standard error. Since, however, an actual distribution of the errors is unavailable, the probable error has no independent status and can be defined only as .67449 of a standard error. It is incorrect to use the term "probable error" for the quartile deviation and unnecessary to use it as $.67499\sigma$ in connection with the dispersion of measures which are not errors.

Measures of Relationship

We have thus far limited our discussion to a single frequency distribution. Frequently, however, two or more series of data have a one-to-one correspondence between their individual measures, such as maximal daily temperatures of New York and corresponding values for Washington, or statures of a group of persons and their respective weights. An interesting question immediately arises as to whether, or how far, are such series related. Does the position of an individual measure in one distribution more or less indicate that of the corresponding measure in the other distribution? Is the relatively tall person relatively heavy? How accurately can one estimate the temperature of New York from the corresponding value for Washington? How much more accurately can one do it with than without the help of the Washington record? Such questions are treated under the subject of *correlation*.

Of course any number of variables may be mutually correlated, but let us first consider the simple case of two quantitative series. The frequencies have to be tabulated by double entry, as each measure must be properly located with respect to two scales. For graphic representation, since two axes are necessary for the two scales, the frequencies are either shown in the third dimension or represented by the densities of dots or other signs each of which denotes an individual measure. The latter method gives the *scatter-diagram,* a few examples of which are shown in Figure 1.

If we now find the mean of each column in the scatter-diagram, we can determine a curve which will best fit these means, by the method of least squares to be treated in the following chapter. The fitted curve will then give the best estimates for the second variable, y, from the first, x. Similarly, a curve can be fitted to the means of the rows and give estimates of x from y. It will be noted that when two series are uncorrelated the fitted curves will be a vertical and a horizontal line which can give no better estimate than the mean of the entire distribution for any measure. In such a case the standard error of estimate is equal to the standard deviation of the distribution, and knowledge of the position of a measure along one scale does not in any way help to locate it along the other scale. But when two series are correlated, the dispersion within each array—a term for both column and row—is less than that of the entire distribution.



NO CORRELATION RECTILINEAR CORRELATION CURVILINEAR CORRELATION

FIGURE 1

and the curve fitted to the means of the arrays will give estimates with a proportionately reduced standard error. As correlation attains perfection, all measures will lie on the fitted curve and the standard error of estimate is reduced to zero.

It will have been obvious from the preceding discussion that there are essentially two important problems in the correlation of quantitative series: to fit curves to the means of the arrays and to determine the standard errors of estimate based upon the fitted curves.

In the majority of cases obtaining in correlated series, a straight line will give a good fit to the means of the columns and to the means of the rows. Of course, one may use the straight line in all situations, but two series which are highly correlated in a curvilinear manner may yield a very much lower rectilinear correlation. We shall later consider a criterion by which it is to be judged whether a given situation has a rectilinear or curvilinear correlation. However, a general inspection of the scatter-diagram will usually give a fairly accurate idea.

RECTILINEAR CORRELATION

In rectilinear correlation the derivation of one single coefficient is sufficient for determining the two best fitting lines and the standard errors of estimate. To simplify our discussion by eliminating the means and standard deviation, let us use the letters z_1 and z_2 to designate the relative positions of a pair of measures in two distributions, and define them by

$$z_1 = \frac{x}{\sigma_1} = \frac{X - M_1}{\sigma_1} \qquad z_2 = \frac{y}{\sigma_2} = \frac{Y - M_2}{\sigma_2} \qquad [19]$$

Such are termed *standard measures* and are independent of the units of measurement. A distribution of standard measures will have a mean of zero and a standard deviation of one. They are very useful in comparing measures in different distributions. In perfect rectilinear correlation, the corresponding z scores in two distributions are exactly equal. If the two variables are negatively or inversely related such that a z score in one distribution is as much above the mean as the corresponding z score in the other is below, then the two scores will be numerically equal but have opposite signs, and the difference between them will be twice the magnitude of either. The square of the difference between corresponding z scores in two distributions, $(z_1 - z_2)^2$, will thus vary between 0 and 4 times z_1^2, or z_2^2, and since the standard deviation of a distribution of standard measures is always one,

$$\frac{\Sigma\,(z_1 - z_2)^2}{N}$$

will vary between 0 and 4, and is clearly a measure of correlation. The coefficient in actual use, known as the *product-moment coefficient* and designated by the letter r, has a range from +1.00 for perfect positive correlation through 0 for no correlation to –1.00 for perfect negative correlation. It is

$$r = 1 - \frac{\Sigma\,(z_1 - z_2)^2}{2N} = \frac{\Sigma\, z_1 z_2}{N} \qquad [20]$$

If we now use r as the ratio of the estimated z_1 to z_2 scores, thus $z_1 = rz_2$, the error of estimate in any single instance is $(z_1 - rz_2)$ and the standard error of estimate, here designated k, is given by the following equation:

$$k^2 = \frac{\Sigma(z_1 - rz_2)^2}{N} = 1 + r^2 - 2r\frac{\Sigma z_1\ z_2}{N} = 1 - r^2 \qquad [21]$$

To prove that r gives the least error of estimate, let us employ $r + e$ in place of r, where e may be any positive or negative quantity, and see if the standard error of estimate is in any way reduced. Accordingly,

$$k^2 = \frac{\Sigma[z_1 - (r + e)\ z_2]^2}{N} = 1 + (r + e)^2 - 2(r + e)r = 1 - r^2 + e^2$$

Since $\Sigma(z_1 - z_2)^2\ /N$ is not less than 0 nor greater than 4, it follows that r^2 is limited by the extreme values of 0 and 1. Consequently $(1 - r^2)$ is always positive and it is obvious that the best value for e is zero. In other words, any deviation from r, as given by [20], would increase the standard error of estimate.

The ratio used in estimating measures in one distribution from the corresponding ones of another is termed the *coefficient of regression*, and we see that r, the coefficient of correlation, is the same as the coefficient of regression when dealing with z scores. If we deal with deviation measures instead of standard measures, then there are two regression coefficients,

$$b_{12} = r\frac{\sigma_1}{\sigma_2} \qquad b_{21} = r\frac{\sigma_2}{\sigma_1} \tag{22}$$

b_{12} for estimating x from y, and b_{21} for estimating y from x. Using letters with a bar to designate estimated scores as distinguished from observed scores, the regression equation may be written in any of the following forms:

$$\bar{z}_1 = rz_2 \qquad \bar{z}_2 = rz_1 \tag{23}$$

$$\bar{x} = r\frac{\sigma_1}{\sigma_2}y = b_{12}\,y \qquad \bar{y} = r\frac{\sigma_2}{\sigma_1}x = b_{21}\,x \tag{23a}$$

$$\bar{X} = r\frac{\sigma_1}{\sigma_2}Y + \left(M_1 - r\frac{\sigma_1}{\sigma_2}M_2\right) = b_{12}\,Y + (M_1 - b_{12}\,M_2)$$

$$\bar{Y} = r\frac{\sigma_2}{\sigma_1}X + \left(M_2 - r\frac{\sigma_2}{\sigma_1}M_1\right) = b_{21}\,X + (M_2 - b_{21}\,M_1) \tag{23b}$$

The value for r can, of course, also be stated in terms of deviation scores in place of standard scores. Thus:

$$r = \frac{\Sigma xy}{N\sigma_1\sigma_2} \tag{20a}$$

Some of the other variations, using gross scores instead of deviation scores, or sums or differences instead of cross products, are given as follows:

$$r = \frac{\Sigma XY - NM_1\,M_2}{N\sigma_1\sigma_2} \tag{20b}$$

$$r = \frac{\Sigma(x+y)^2 - \Sigma x^2 - \Sigma y^2}{2\,N\,\sigma_1\sigma_2} \tag{20c}$$

$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma(x-y)^2}{2\,N\,\sigma_1\sigma_2} \tag{20d}$$

The standard error of estimate, as given by $k = \sqrt{1-r^2}$, is termed the *coefficient of alienation* and applies only in case of standard measures, the standard deviation of which is always one. It is more exactly the ratio of the standard error of estimate to the standard deviation of the distribution. When the standard deviation of a distribution differs from one,

$$\sigma_{1\cdot2} = \sigma_1 k = \sigma_1\sqrt{1 - r^2} \qquad [24]$$

where $\sigma_{1\cdot2}$ designates the standard error of estimating x from y. The corresponding standard error of estimating y from x is similarly

$$\sigma_{2\cdot1} = \sigma_2 k = \sigma_2\sqrt{1 - r^2} \qquad [24a]$$

The relation between the coefficient of correlation on one hand and the alienation coefficient, or the ratio of the standard error of estimate to the standard deviation of the distribution, on the other,

$$r = \sqrt{1 - k^2} = \sqrt{1 - \frac{\sigma_{1\cdot2}^2}{\sigma_1}} = \sqrt{1 - \frac{\sigma_{2\cdot1}^2}{\sigma_2^2}} \qquad [21a]$$

gives a very significant as well as accurate interpretation of the correlation coefficient. Persons making use of the coefficient should be warned that it reveals conditions which, though necessary, are insufficient for inferences as to causal relations and common elements and other hypotheses involving extra-statistical considerations.

From equation [21*a*] it will be seen that the constancy of r depends upon the constancy of k. As the range of a distribution increases, however, errors of estimate remain relatively constant, and consequently r will increase with the standard deviation. Assuming, then, that errors of estimate are equal for different ranges, we have the following relation between the dispersion of the distributions and the magnitude of the correlation coefficient:[1]

$$\frac{\sigma}{\Sigma} = \frac{\sqrt{1 - R}}{\sqrt{1 - r}} \qquad [25]$$

where r is the correlation coefficient derived from distributions with a standard deviation of σ, and R from distributions with a standard deviation of Σ. Since the range of data is more or less arbitrary and accidental in any given study, it should always be taken into account in interpreting the correlation coefficient derived.

The standard error of the product-moment coefficient of correlation is given by

$$\sigma_r = \frac{k^2}{\sqrt{N}} = \frac{1 - r^2}{\sqrt{N}} \qquad [26]$$

When N, the population, is small, the formula is inexact and tends toward underestimation. If the population is very small, say less than 10, the standard error will be so large that the correlation coefficient is practically meaningless.

---

[1] We may expect the conditions leading to equation [25] to commonly be met by reliability coefficients, but not by other correlation coefficients.

If measures of two series to be correlated are expressed in consecutive ranks instead of graded scores, a special formula, due to Spearman, is available which will considerably simplify the procedure. It is derived by substituting $\sqrt{N^2-1} / \sqrt{12}$, the standard deviation of N ranks, for both σ_1 and σ_2 in formula [20d], yielding

$$\rho = 1 - \frac{6\,\Sigma d^2}{N(N^2-1)} \qquad [27]$$

where d designates the difference between corresponding ranks in the two series. The value of ρ is exactly equal to r upon the assumption that the consecutive ranks are equidistant. A slight error is introduced in case of occasional ties in rank. Pearson has shown that if ρ is calculated from ranks which are forced upon normally distributed measures, its discrepancy from r, obtained directly from the scores, is given by the following formula:

$$r = 2 \sin \frac{\pi}{6} \rho \qquad [28]$$

This correction, however, is relatively small, and is not generally useful since lack of normality of distribution may introduce errors larger in magnitude. While ρ is more easily calculated than r, there is always some loss of accuracy from warping graded scores into ranks. The standard error of ρ, as derived by Pearson, is about five per cent larger than that of r.

The Correlation Ratio

We have already seen that one of the definitions of r is $\sqrt{1-\sigma_{1.2}^2 / \sigma_1^2}$ where $\sigma_{1.2}^2$ is the mean square error of estimate or mean square deviation of the measures from the regression line. If the means of the arrays lie on the regression line, then $\sigma_{1.2}^2$ is at the same time the average of the mean square deviations within each of the arrays, and the correlation coefficient gives the true magnitude of relationship between the two variables. Sometimes, however, a straight line does not give a good fit to the means of the arrays, and the measures within an array will deviate significantly more from the regression line than from the mean of the array. In other words, the standard error of estimate will be significantly reduced if in place of the fitted straight line we use the means of the arrays, to which a curve other than the straight line can be better fitted. The method of curve-fitting, which will be described in the following chapter, can readily be applied, if a given problem of correlation justifies the labor. But the method of the *correlation ratio,* using the mean square deviation from the means of the arrays instead of from the straight line of regression, will show the maximum possible reduction of the standard error of estimate. There are two correlation ratios for each scatter-diagram:

$$\eta_{12}=\sqrt{1-\frac{\sigma_{ax}^2}{\sigma_1^2}} \qquad \eta_{21}=\sqrt{1-\frac{\sigma_{ay}^2}{\sigma_2^2}}$$

where σ_{ax}^2 designates the average mean square deviation within each of the x arrays and σ_{ay}^2 designates the average mean square deviation within each of the y arrays. These values can be directly calculated from the data, but considerable labor will be saved if we make use of the relation, already expressed in formula [5], that the mean square deviation of measures within any array from the mean of the entire distribution is equal to the mean square deviation of the same measures from the mean of the array plus the square of the deviation of the mean of the array from the mean of the distribution. Thus:

$$\left.\begin{aligned}\eta_{12}&=\sqrt{1-\frac{\sigma_{ax}^2}{\sigma_1^2}}=\sqrt{\frac{\sigma_1^2-\sigma_{ax}^2}{\sigma_1^2}}=\sqrt{\frac{\sigma_{mx}^2}{\sigma_1^2}}=\frac{\sigma_{mx}}{\sigma_1}\\ \eta_{21}&=\sqrt{1-\frac{\sigma_{ay}^2}{\sigma_2^2}}=\sqrt{\frac{\sigma_2^2-\sigma_{ay}^2}{\sigma_2^2}}=\sqrt{\frac{\sigma_{my}^2}{\sigma_2^2}}=\frac{\sigma_{my}}{\sigma_2}\end{aligned}\right\} \quad [29]$$

where σ_{mx} and σ_{my} designate the standard deviations of the means of the x arrays and y arrays, respectively, from the mean of the distribution. In calculating these standard deviations, care should be taken to weight the different arrays according to the number of measures in each.

The derivation of η does not require that the successive arrays should be equidistant. In fact, the arrays need not be in any order, they may be discontinuous categories. We can therefore calculate a correlation ratio between two series of data of which only one is quantitative.

That the magnitude of η is in part a function of the coarseness with which data are grouped may already have been noticed. If the classes are so minute that but a single measure is found in any array or category, η would of necessity have the maximum value of unity and would of course have no real significance. This extreme hypothetical case is not likely to be found in an actual situation, but it illustrates how too fine grouping tends to dilate η beyond the true relationship between the variables. On the other hand, if a series is coarsely divided into very few classes, the dispersion within each array will remain relatively large and consequently it will yield a value for η which is too small. Errors of underestimate due to coarse grouping are common to both r and η, but errors or overestimate due to fine grouping are peculiar to the correlation ratio. The correction for fine grouping is usually the more important and is given by

$$ {}_{t}\eta^2 = \frac{\eta^2 - \dfrac{k-3}{N}}{1 - \dfrac{k-3}{N}} \qquad [30] $$

${}_{t}\eta$ here designating the most probable true ratio and k standing for the number of categories or arrays. This formula is to be used when η^2 is definitely larger than $(k-1)/N$.

The correlation ratio is always positive, and can neither be greater than one nor less than zero. Its numerical value will be equal to the correlation coefficient derived from the same data if the means of the arrays lie strictly on a straight line. But mere chance fluctuations in the means of the arrays will obscure true rectilinearity and result in yielding a correlation coefficient lower than the correlation ratio. In order, therefore, to test for rectilinearity, the function,

$$ \zeta = \eta^2 - r^2 \qquad [31] $$

should be expressed in terms of its standard error, which Blakeman has given as follows:

$$ \sigma_{\zeta} = \frac{2}{\sqrt{N}} \left\{ \zeta[(1 - \eta^2)^2 - (1 - r^2)^2 + 1] \right\}^{\frac{1}{2}} \qquad [32] $$

Formula [32] was derived before formula [30] was available, but it seems reasonable to apply it to the corrected, rather than the raw, η. The standard error of η is usually taken as

$$ \sigma_{\eta} = \frac{1 - \eta^2}{\sqrt{N}} \qquad [33] $$

which, however, does not take into consideration errors due to extreme grouping and consequently may yield a value which is too small.

The Method of Contingency

The correlation coefficient is applicable only to quantitative series, or to qualitative series which are ordered and are capable of transformation into quantitative ones. The correlation ratio has a wider range of application in that one of the two series does not need to be quantitatively graded. But for strictly qualitative data, some different method of treatment is necessary. The fundamental proposition underlying the method of contingency is that the probability of two independent events jointly occurring is equal to the product of the probability of the first event multiplied by that of the second. Thus if successive tosses of a coin are independent and if the probability of heads' turning up any one time is 1/2, then the probability of heads' turning up twice in succession is 1/2 times 1/2 or 1/4. Similarly, if imbecility and deaf-mutism are uncorrelated, and if we find in a given population one imbecile in every 600 people and one deaf-mute in every 2000, then we expect to find one person who is both imbecile and deaf-mute in about 1,200,000. But suppose the proportion of deaf-mute imbeciles actually found in a population is, say, as high as one in every 72,000, then we may conclude that there is a positive association between imbecility and deaf-mutism. We thus make use of the divergence in the frequency of a category from what is expected upon the assumption of independence as a basis for the measurement of correlation between qualitative series.

Let us use n_s to designate the frequency in a certain category, say, the profession of teaching, n_t to designate the frequency in a category of a different classification, say, death caused by diphtheria, n_{st} to designate the number of those in the teaching profession who died of diphtheria, and N to designate the entire sample of population under consideration. The condition of independence between teaching as a profession and diphtheria as a cause of death is then expressed by

$$\frac{n_{st}}{N} = \frac{n_s}{N} \cdot \frac{n_t}{N} \quad \text{or} \quad n_{st} = \frac{n_s \, n_t}{N}$$

The divergence from this condition,

$$d_{st} = n_{st} - \frac{n_s \, n_t}{N} \qquad [34]$$

when squared and divided by $(n_s \, n_t/N)$, the frequency expected upon the assumption of independence, is a measure of association between the two attributes in question. But when we need a measure of general relationship between profession and cause of death, we must also consider the other professions and the other causes of death, and give them similar treatment. The function,

$$\chi^2 = \Sigma \left[\frac{d^2_{st}}{\frac{n_s \, n_t}{N}} \right] \qquad [35]$$

called the *square contingency*, is thus a summation of as many items as the number of professions multiplied by that of causes of death. The *mean square contingency*,

$$\phi^2 = \frac{\chi^2}{N} \qquad [36]$$

is equivalent to r^2 in the case of a 2x2 categorical distribution where the frequencies are all concentrated at the four points. For manifold classifications, the measure generally in use is neither χ nor ϕ, but the *coefficient of contingency,*

$$C = \sqrt{\frac{\phi^2}{1+\phi^2}} = \sqrt{\frac{\chi^2}{N+\chi^2}} \qquad [37]$$

As will be seen, C is zero when χ^2 is zero and approaches one as χ^2 increases. But it should be noticed that C will not reach one unless χ^2 is infinite. Yule (51) has in fact shown that C has an upper limiting value much less than one if the number of categories is small. Thus, for example, for a 2x2 classification of 4 categories C cannot exceed .707, and for a 5x5 classification of 25 categories C cannot exceed .894. The same writer has therefore recommended to restrict the use of C to 5x5 or finer classifications. Pearson (25) has derived certain formulae for corrections. But the errors are small if the number of categories is 5x5 or more, or if C is low.

The standard error of C is given by

$$\sigma_C = \frac{1}{\sqrt{N}} \sqrt{\frac{\frac{\psi^3}{\phi^2} + 1 - \phi^2}{(1+\phi^2)^3}} \qquad [38]$$

where ϕ^2 has already been defined by [36] and where

$$\psi^3 = \frac{1}{N} \Sigma \left[\frac{d^3{}_{st}}{\left(\frac{n_s\, n_t}{N}\right)^2} \right]. \qquad [39]$$

CHAPTER 23

THE STATISTICAL TREATMENT OF CERTAIN TYPICAL PROBLEMS

TRUMAN L. KELLEY
Stanford University

AND

EUGENE SHEN
China Institute in America

PARTIAL CORRELATION

All measures of correlation, whether the series are quantitative or qualitative, whether the regression is rectilinear or curvilinear, may be applied either to data which are found in the universe at large or to data which are limited by certain factors. Thus we may study the correlation between height and weight among individuals of a constant age as well as among human beings in general; we may calculate the coefficient of contingency between the racial stock and occupation of immigrants at large in the United States or limit the study to immigrants of a certain educational status. The measures of relationship for selected data, known as measures of *partial correlation,* can, of course, be directly calculated in the same manner as for unselected data. But our problem here is to derive such measures indirectly. We shall restrict our treatment to partial correlation of rectilinear regressions, though the concept is common to other situations.

The simplest case of partial correlation involves three variables. Such data yield the correlation coefficients, r_{12}, r_{13}, and r_{23}, which, to be distinguished from the partial correlation coefficients, we shall call total correlation coefficients. From these three total correlation coefficients we can derive three partial correlation coefficients, as follows:

$$\left.\begin{aligned} r_{12\cdot3} &= \frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r^2_{13})\ (1-r^2_{23})}} \\ r_{13\cdot2} &= \frac{r_{13}-r_{12}r_{23}}{\sqrt{(1-r^2_{12})\ (1-r^2_{23})}} \\ r_{23\cdot1} &= \frac{r_{23}-r_{12}r_{13}}{\sqrt{(1-r^2_{12})\ (1-r^2_{13})}} \end{aligned}\right\} \qquad [40]$$

If we have as the three variables the stature of fathers, mothers, and sons, respectively, then r_{12} is the correlation between fathers and mothers, while $r_{12 \cdot 3}$ is that between fathers and mothers in case the sons are of the same stature. Similarly, $r_{13 \cdot 2}$ is the correlation between fathers and sons in case the mothers are equal in stature, and $r_{23 \cdot 1}$ is the correlation between mothers and sons with the stature of fathers constant.

Usually the correlation coefficient between two variables is lowered when a third variable is "held constant" or "partialed out." But it is not necessarily the case. The correlation, for instance, between the area of a rectangle and one of the two dimensions will be raised from a moderately high coefficient to unity when the other dimension is held constant. In this case we may say that by partialing out the influence of a second dimension the true relationship between the area of a rectangle and one of the two dimensions is revealed, for whatever correlation is found between the two dimensions of a rectangle is not intrinsic but accidental, and thus by holding one of the two dimensions constant a disturbing factor is eliminated.

Inferences of this kind, however, must be made only with extreme caution. In general, no causal relationship can be deduced from partial correlation coefficients any more than from total correlation coefficients. When one speaks of the partial correlation between the stature of father and son with that of mother constant as showing the true influence of the father isolated from that of the mother, it is true only in so far as we know that both the father and mother have an influence on the son and that the two are not intrinsically correlated. That the premises lie quite beyond the partial correlation coefficient is obvious as soon as one considers a different phase of the same situation. It would be highly absurd to interpret "holding the stature of sons constant" in correlating the stature between father and mother as "eliminating the disturbing influence of the son."

Analogous to partial correlation coefficients of the first order where a third variable is held constant, we have partial correlation coefficients of the second and higher orders where two or more variables are held constant. Thus $r_{12 \cdot 34}$ designates the partial correlation coefficient between the first two variables with the third and fourth constant. The following formulae, which are equivalent and serve as a check upon each other, show how such coefficients are derived:

$$\left.\begin{aligned} r_{12 \cdot 34} &= \frac{r_{12 \cdot 3} - r_{14 \cdot 3} r_{24 \cdot 3}}{\sqrt{(1 - r^2_{14 \cdot 3})\ (1 - r^2_{24 \cdot 3})}} \\ r_{12 \cdot 43} &= \frac{r_{12 \cdot 4} - r_{13 \cdot 4} r_{23 \cdot 4}}{\sqrt{(1 - r^2_{13 \cdot 4})\ (1 - r^2_{23 \cdot 4})}} \end{aligned}\right\} \qquad [41]$$

One important difference must be noted between the indirectly derived partial correlation coefficients and correlation coefficients directly calculated from selected data. Suppose we are studying the correlation between height and weight and wish to hold age constant. If we calculate a separate coefficient for each age, it is obvious that they need not be all equal, and the variation may be due to significant causes worthy of further investigation. But the partial correlation coefficient gives a single value for the various ages and can never reveal such differences.

MULTIPLE CORRELATION

The problem of *multiple correlation* is the estimation, with minimal error, of a variable from several others. The variable to be estimated is called the *dependent* variable and all the others *independent* variables, the terms "dependent" and "independent" here having no other implication than that they are temporarily so considered in the problem. More specifically, then, we desire (1) a regression equation as a means of estimating the dependent variable knowing the independent variables, and (2) a correlation coefficient showing the strength of relationship between the dependent variable and an optimal combination of the independent variables. Confining our treatment to the simple case of three variables, and using the zero subscript for the dependent variable, we can write down the regression equation of deviations from the means as

$$\bar{x}_0 = b_{01\cdot 2}x_1 + b_{02\cdot 1}x_2 \qquad [42]$$

If x_0 represents a measure of leadership, x_1 a measure of adaptability, and x_2 a measure of persistence, $b_{01\cdot 2}$ is the regression of that part of leadership which is independent of persistence upon that part of adaptability which is independent of persistence, and $b_{02\cdot 1}$ is similarly the regression of that part of leadership which is independent of adaptability upon that part of persistence which is independent of adaptability. The values for these *partial regression coefficients* are given by the following formulae:

$$\left.\begin{aligned} b_{01\cdot 2} &= r_{01\cdot 2}\frac{\sigma_{0\cdot 2}}{\sigma_{1\cdot 2}} = \frac{\sigma_0}{\sigma_1}\left[\frac{r_{01} - r_{02}r_{12}}{1 - r^2_{12}}\right] \\ b_{02\cdot 1} &= r_{02\cdot 1}\frac{\sigma_{0\cdot 1}}{\sigma_{2\cdot 1}} = \frac{\sigma_0}{\sigma_2}\left[\frac{r_{02} - r_{01}r_{12}}{1 - r^2_{12}}\right] \end{aligned}\right\} \qquad [43]$$

It will be seen that the partial regression coefficients are analogous to the total regression coefficients. The total regression of x_0 upon x_1, is $b_{01} = r_{01}\ \sigma_0/\sigma_1$. But here, for $b_{01\cdot 2}$, we are considering only parts of x_0 and x_1, which are independent of x_2, so we have $r_{01\cdot 2}$ in place of r_{01}, $\sigma_{0\cdot 2}$ in place of σ_0, and $\sigma_{1\cdot 2}$ in place of σ_1.

The coefficients $b_{01\cdot 2}$ and $b_{02\cdot 1}$ are the weights for x_1 and x_2, respectively, in estimating x_0. The standard error of estimate,

$$\sigma_{0\cdot 12} = \sqrt{\frac{\Sigma(x_0 - \bar{x}_0)^2}{N}}$$

is equal to σ_0, the standard deviation of the x_0 measures, multiplied by the alienation coefficient,

$$k_{0\cdot12} = \frac{1}{k_{12}}\sqrt{1-r^2_{01}-r^2_{02}-r^2_{12}+2r_{01}r_{02}r_{12}} \qquad [44]$$

where k_{12} is equal to $\sqrt{1-r^2_{12}}$, as already defined in formula [21]. The value of the terms under the radical is never greater than one nor less than zero, if the various r's are derived from a consistent set of data.

The multiple correlation coefficient, $r_{0\cdot12}$, is the correlation between x_0 and $\bar{x}_0$ and can be so calculated, but it is more simply derived from $k_{0\cdot12}$ or $\sigma_{0\cdot12}$, as follows:

$$r_{0\cdot12} = \sqrt{1-k^2_{0\cdot12}} = \sqrt{1-\frac{\sigma^2_{0\cdot12}}{\sigma^2_0}} \qquad [45]$$

For problems involving more than three variables, or for the general problems of n variables, the various formulae are best written in terms of determinants. Let

$$\Delta = \begin{vmatrix} 1 & r_{01} & r_{02} & \cdots & r_{0n} \\ r_{01} & 1 & r_{12} & \cdots & r_{1n} \\ r_{02} & r_{12} & 1 & \cdots & r_{2n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ r_{0n} & r_{1n} & r_{2n} & \cdots & 1 \end{vmatrix} \qquad [46]$$

and let Δ_{pq} designate a determinant derived from Δ by crossing out the row having p as one of the subscripts of the r's throughout and the column having q as one of the subscripts throughout. Then

$$k_{0\cdot12\ldots n} = \sqrt{\frac{\Delta}{\Delta_{00}}} \qquad [47]$$

$$r_{0\cdot12\ldots n} = \sqrt{1-\frac{\Delta}{\Delta_{00}}} \qquad [48]$$

$$b_{0p\cdot12\ldots n} = \frac{\sigma_0}{\sigma_p}\left(\frac{-(-1)^p\Delta_{0p}}{\Delta_{00}}\right) \qquad [49]$$

In [49] the subscripts beyond the point, namely, $_{12\ldots n}$, do not contain the subscript p. Note that in the formula for the partial regression coefficient the determinant in the numerator is positive when p is odd and negative when p is even, thus:

$$b_{01\cdot23\cdots n} = \frac{\sigma_0}{\sigma_1}\left(\frac{\Delta_{01}}{\Delta_{00}}\right) \qquad \text{while} \qquad b_{02\cdot13\cdots n} = \frac{\sigma_0}{\sigma_2}\left(\frac{-\Delta_{02}}{\Delta_{00}}\right)$$

The calculation of a regression coefficient for three variables is readily accomplished by formula [43], by Kelley's *Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression* (10),[4] by Kelley's *Alignment Chart* (11), and by sundry slightly different arrangements for performing the arithmetic steps. For four, five, or six variables there is available a layout made by Yule (51), another by Kelley (11), and also the Doolittle method (22). For a still larger number of variables one may use the Doolittle method or an iteration method especially designed by Kelley and Salisbury for the many variable problem (19, 30, 48, 18).

The standard errors of multiple correlation coefficients and partial correlation coefficients are similar to that of the total correlation coefficients between two variables. Thus:

$$\sigma_{r_{0\cdot12\cdots n}} = \frac{k^2_{0\cdot12\cdots n}}{\sqrt{N}} = \frac{1-r^2_{0\cdot12\cdots n}}{\sqrt{N}} \qquad [50]$$

$$\sigma_{r_{01\cdot23\cdots n}} = \frac{k^2_{01\cdot23\cdots n}}{\sqrt{N}} = \frac{1-r^2_{01\cdot23\cdots n}}{\sqrt{N}} \qquad [51]$$

These are only approximate formulae, and not to be used if n, the number of independent variables, is at all large in comparison with N, the size of the population (4).

If the number of variables is equal to the population, any of them which is considered the dependent variable will of necessity have a perfect multiple correlation with the remaining, independent variables, provided of course that none of the latter is redundant or has a perfect correlation with another. This proposition is easily demonstrated, for if we write down a regression equation of the following form for each individual,

$$x_0 = b_{01\cdot23\ldots}X_1 + b_{02\cdot13\ldots}x_2 + \ldots + b_{0n\cdot12\ldots}X_n + C$$

there will be just as many such equations as required for the solution of the regression coefficients and the constant.

CURVE-FITTING

If the number of observations is sufficiently large in a study, it is often desirable to find an algebraic expression for a curve which will best fit the data. The procedure involves three distinct steps: the type of curve must first be selected according to the general character of data, an exact equation representing the selected type of curve can then be determined by certain numerical derivatives, and finally the curve or equation may be tested for goodness of fit. While the last two steps are mechanical processes, a large degree of freedom in the first is left to the judgment of the investigator. He has to decide whether he wants to fit a straight line, a parabola, an exponential function, or some other type, before he can determine the constants of the equation and find its divergence from the data. It is a mathematical principle that by selecting a function which contains as many constants as the number of points plotted from a set of data, a curve can always be determined which will pass through all the points. Such, however, is far from the most desirable procedure, as it would take into account all the chance irregularities and spuriously complicate the equation. The simpler curve is always preferred, if its deviation from the data can reasonably be attributed to errors of observation and sampling.

---

[4] Now out of print, available in mimeograph form from the Bureau of Administrative Research, University of Cincinnati, Ohio.

For frequency distributions, Karl Pearson has developed a method of fitting curves by the use of moments. It is obvious that the first and second moments, or the mean and the standard deviation, are altogether inadequate for the purpose, since they do not in any way determine the type of a curve aside from its reference to a scale. The necessary additional information is furnished by the beta coefficients:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad [52]$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \qquad [53]$$

The μ's have already been defined in formula [4] in the preceding chapter, but it should be noted that μ_4, like μ_2, has an error due to grouping, and Sheppard's correction should also be applied. Corresponding to a difference of 1/12 of an interval squared in μ_2, μ_4 is too large by $(\mu_2/2) + (1/80)$ of an interval to the fourth power, μ_2 here having already been corrected. No Sheppard's corrections on account of grouping are required in μ_3.

The first of these beta's is a coefficient of *skewness,* and is equal to zero if a distribution is symmetrical, since the third powers of the deviations in the two directions from the mean will then just balance each other. The magnitude of the coefficient is in proportion to the variation of the distribution from perfect symmetry. Skew distributions are rather typical of certain data, such as reactions which attempt to be as short as possible, intelligence ratings of 12-year-olds in the fourth grade where only the relatively dull remain, and test scores of a group for whom the test is too easy or too difficult to make an adequate scale.

The coefficients of *kurtosis,* β_2, a function of the second and fourth moments, has a numerical value of at least 1.00, a limit reached when a distribution is an even dichotomy, i.e., composed of two categories of equal population. It approaches 3.00 as a frequency distribution approaches *mesokurtosis,* which is the kurtosis of a normal distribution. If it is less than 3.00, the distribution is termed *platykurtic,* with a flat top and contracted ends, tending toward a rectangular shape. As it increases from 3.00, the distribution becomes *leptokurtic,* with a high peak steeply descending to long tails.

Skewness and kurtosis are sufficient for describing practically all types of unimodal distribution curves. Moments higher than the fourth are generally so unreliable that their inclusion in determining the type of a distribution curve would increase the labor without increasing the accuracy of the results. After obtaining the beta constants from the data, therefore, the type of the curve is practically fixed. It now remains to derive the constants of the equation. The simplest functions have only one constant to determine, in addition to N, the population. Some have two constants, others have as many as four, depending upon the type of curve chosen.

It will be remembered that the normal distribution has a β_1 equal to 0, and a β_2 equal to 3. In the equation of such distributions with the mean as origin,

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} \qquad [54]$$

the only constant required besides N is σ, the standard deviation of the distribution. There are a large number of obtained distributions to which the normal curve will give a good fit, but if a distribution is definitely skew, or deviates markedly from mesokurtosis, it is necessary to select another type. There are altogether 15 different types of curves as given by Pearson. It is beyond the scope of the present chapter even to give a brief description, and the reader interested in the subject must consult the references cited at the end of the chapter.

Curves other than those of frequency distributions, such as growth curves and learning curves, have not been systematically classified according to any constants comparable to the beta coefficients. Consequently, there is no definite procedure to be followed in selecting the type of curve which one may fit to a set of data. For purposes of illustration, we shall consider the nth-order parabola, which is a very general and very useful function. The general equation is written as follows:

$$\overline{Y} = C_0 + C_1X + C_2X^2 + \ldots + C_nX^n \qquad [55]$$

where $\overline{Y}$ represents the calculated ordinates as distinguished from the observed ordinates to be designated Y, where the C's are constants to be determined from the data, and where the number of terms to be included can be determined only upon past experience and experimental trial. If the addition of a term does not appreciably reduce the divergence of a fitted curve from obtained data, the simpler curve is always to be preferred.

A method entitled to special confidence for determining the constants, when the number of terms has been decided, is the *method of least squares,* i.e., the constants are to be chosen such that, within the conditions stipulated by the type of curve selected, the squares of the errors will add to a minimum. In other words, we want to make u a minimum, where

$$u = \Sigma(Y - \bar{Y})^2 = \Sigma[Y - (C_0 + C_1X + C_2X^2 + \ldots + C_nX^n)]^2 \quad [56]$$

It is obvious that when a curve is at its minimum, its slope is zero, or just passes from a negative to a positive value. And the slope of a curve is given by its first derivative. Accordingly, to make u a minimum, consider it as a function of C_0, C_1, etc., and set to zero the partial derivatives of u with respect to each of the C's and thereby derive the following $n + 1$ *normal equations* for the solution of the $n + 1$ C's:

$$\left.\begin{array}{l}
\Sigma Y = C_0\Sigma 1 + C_1\Sigma X + C_2\Sigma X^2 + \ldots + C_n\Sigma X^n \\
\Sigma X Y = C_0\Sigma X + C_1\Sigma X^2 + C_2\Sigma X^3 + \ldots + C_n\Sigma X^{n+1} \\
\Sigma X^2 Y = C_0\Sigma X^2 + C_1\Sigma X^3 + C_2\Sigma X^4 + \ldots + C_n\Sigma X^{n+2} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\Sigma X^n Y = C_0\Sigma X^n + C_1\Sigma X^{n+1} + C_2\Sigma X^{n+2} + \ldots + C_n\Sigma X^{2n}
\end{array}\right\} \quad [57]$$

In case the various Y's are each derived from a set of observations, they should be weighted according to the number of observations from which each is derived. This would alter the preceding equations by adding to each term its appropriate weight proportional to the frequency of observations.

As an illustration of the procedure, we shall fit a curve to Shirokogoroff's data (33) on the stature of school boys in Shanghai. On account of the limited observations—295 cases ranging between the ages 8 and 19—the fitted curve can hardly be expected to be worthy of the labor involved. But, for our purpose, it has the advantage, in addition to showing the mathematical procedure, of revealing the limitations of the method. The original data are given in the first three columns of the accompanying table, and the remaining columns give calculations necessary to the fitting of a cubic. They are, of course, more than sufficient for fitting a straight line or a second-order parabola. For comparison, we shall derive all the three curves. For a straight line, we have the following pair of normal equations:

$$\begin{array}{l}
445{,}968 = 295\ C_0 + 221\ C_1 \\
413{,}692 = 221\ C_0 + 1821\ C_1
\end{array}$$

Solving for C_0 and C_1, we obtain 1475.7 and 48.08, respectively. Thus the equation for the straight line best fitting the data is

$$\bar{Y} = 1475.7 + 48.08\ X$$

X, it should be noted, is taken as age minus 13.5 years. The equation for the best fitting second-degree parabola is

$$\bar{Y} = 1482.3 + 49.28\ X - 1.213\ X^2$$

which follows from the following set of normal equations:

TABLE 1

Age	f	Y Stature in mm.	X Age-13.5	fY	fXY	fX^2Y	fX^3Y	fX	fX^2	fX^3	fX^4	fX^5	fX^6
8.5	7	1255	−5	8785	− 43925	219625	−1098125	− 35	175	− 875	4375	−21875	109375
9.5	6	1285	−4	7710	− 30840	123360	− 493440	− 24	96	− 384	1536	− 6144	24576
10.5	15	1334	−3	20010	− 60030	180090	− 540270	− 45	135	− 405	1215	− 3645	10935
11.5	20	1349	−2	26980	− 53960	107920	− 215840	− 40	80	− 160	320	− 640	1280
12.5	35	1398	−1	48930	− 48930	48930	− 48930	− 35	35	− 35	35	− 35	35
13.5	58	1487	0	86246	−237685		−2396605	−179		−1859		−32339	
14.5	37	1523	1	56351	56351	56351	56351	37	37	37	37	37	37
15.5	48	1609	2	77232	154464	308928	617856	96	192	384	768	1536	3072
16.5	24	1636	3	39264	117792	353376	1060128	72	216	648	1944	5832	17496
17.5	30	1651	4	49530	198120	792480	3169920	120	480	1920	7680	30720	122880
18.5	15	1662	5	24930	124650	623250	3116250	75	375	1875	9375	46875	234375
					651377		8020505	400		4864		85000	
					−237685		−2396605	−179		−1859		−32339	
Σ	295			445968	413692	2814310	5623900	221	1821	3005	27285	52661	524061

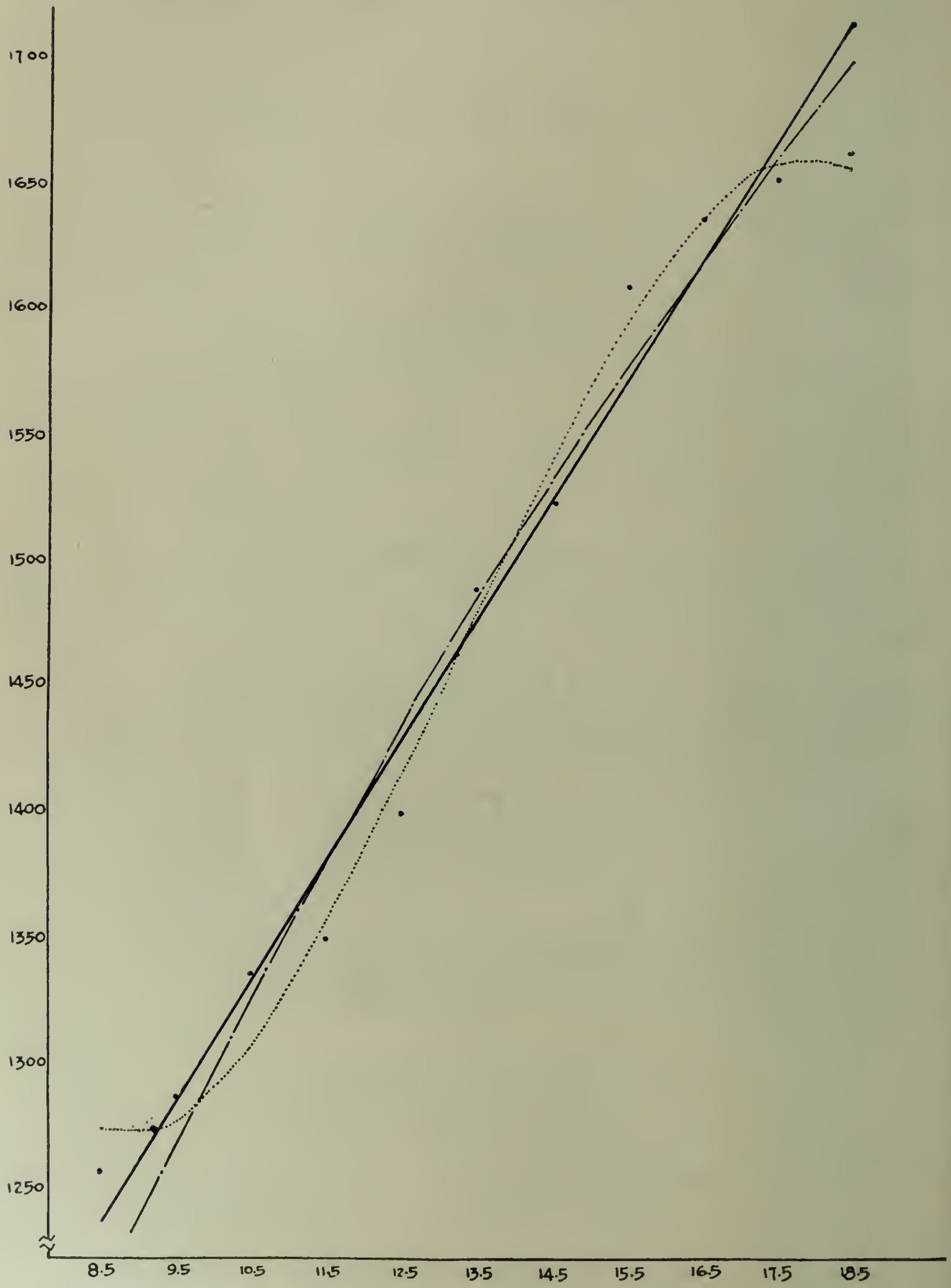

1700
1650
1600
1550
1500
1450
1400
1350
1300
1250
8.5
9.5
10.5
11.5
12.5
13.5
14.5
15.5
16.5
17.5
18.5

FIGURE 2

$$\begin{aligned}445{,}968 &= 295\,C_0 + 221\,C_1 + 1821\,C_2\\ 413{,}692 &= 221\,C_0 + 1821\,C_1 + 3005\,C_2\\ 2{,}814{,}310 &= 1821\,C_0 + 3005\,C_1 + 27285\,C_2\end{aligned}$$

Similarly, the four normal equations,

$$\begin{aligned}445{,}968 &= 295\,C_0 + 221\,C_1 + 1{,}821\,C_2 + 3{,}005\,C_3\\ 413{,}692 &= 221\,C_0 + 1{,}821\,C_1 + 3{,}005\,C_2 + 27{,}285\,C_3\\ 2{,}814{,}310 &= 1{,}821\,C_0 + 3{,}005\,C_1 + 27{,}285\,C_2 + 52{,}661\,C_3\\ 5{,}623{,}900 &= 3{,}005\,C_0 + 27{,}285\,C_1 + 52{,}661\,C_2 + 524{,}061\,C_3\end{aligned}$$

yield the following equation for the best fitting cubic:

$$\overline{Y} = 1477.6 + 64.08\,X - .5537\,X^2 - 1.0220\,X^3$$

Three curves are plotted in Figure 2. As to be expected, the cubic fits the data better than the second-order parabola which is, in turn, better than the straight line. It has been calculated that $\Sigma(Y-\overline{Y})^2$ or $\Sigma f(Y-\overline{Y})^2$ is equal to 184,387 in the case of the straight line, 163,471 for the second-order parabola, and 49,648 for the cubic. The improvement of the cubic over the second-order parabola is thus considerably more than that of the latter over the straight line. But the fitted cubic is really an unreasonable curve, if one extends it beyond the range of the observed data, for there is a sharp inflection at either end of the range. It gives the absurd values, for instance, of 1803 mm. or 71 inches for age 3.5, and 1041 mm. or 41 inches for age 23.5. One should therefore be extremely careful in fitting curves for extrapolation.

However, just recognizing the limitations of the curve which we have fitted is by no means the most thoroughgoing point of view. We have noted that the fitted cubic gives absurd values beyond the range of the observed data. Why absurd? Obviously because they are irreconcilable to our knowledge of the facts. In other words, our general knowledge relative to the growth of stature extends beyond the data made use of in fitting the curve. Why not make use of that general knowledge? Why not, in addition to the data tabulated, put in, say, five cases of 650 mm. for age 0.5 and a case of 1700 mm. for each of the ages from 28.5 to 48.5? To be sure, the resulting curve will not fit the tabulated data so closely, but it will represent the growth of stature much more accurately. Short of actual measurements extending to a wider range, the suggested procedure is a very reasonable course which can be taken to improve the curve.

Another consideration in this connection is the type of curve chosen. We selected the *n*th-order parabola for our illustration primarily because of its flexibility in the number of constants. It is probably not the best type for growth curves. Had we used the logarithmic function,

$$\overline{Y} = a + bX + c\log X \qquad [58]$$

or the hyperbola,

$$\overline{Y} = \frac{X}{a + bX} + c \qquad [59]$$

we would not have obtained such unreasonable results as the cubic has given us, even though the range of the data was so limited.

Psychophysical Methods

The general problem in psychophysics is to determine the *threshold* or *limen,* i.e., the magnitude of stimulus which just brings a sensation to consciousness or which corresponds to a least sense step from some positive point upon the sense scale. The magnitude is accordingly either a *stimulus limen* (*RL*) or a *difference limen* (*DL*). Examples of the two varieties of limen are the lowest audible tone and the least perceptible difference in lifted weights. The various experimental methods in use in psychophysics may be placed in three general classes:

1) The *method of limits,* also variously called the *method of just-noticeable differences, of least differences,* or *of minimal changes,* where a stimulus at first equal to a standard is successively changed until it is noticed by the observer to be just different from the other.

2) The *method of average error,* also called the *method of reproduction* or *of adjustment of equivalent r,* where the observer is to adjust a second stimulus so as to make it equivalent to a standard stimulus presented to him. A modification of this method is the *method of equivalents* in which the two stimuli are applied to different parts of a receptor or to different receptors.

The *method of equal sense distances, of mean gradations, of supraliminal differences,* or *of equal-appearing intervals,* is not strictly different from the two foregoing methods. Here the observer is given two stimuli, r_1 and r_3, to determine a third stimulus, r_2, such that the difference between r_1 and r_2 will be noticed as equal to that between r_2 and r_3. In interpolating r_2, one may either follow the principle of the method of limits or that of the method of average error.

3) The *method of constant stimuli* or the *method of right and wrong cases,* where the observer is to pass judgments upon a series of stimuli with reference to a certain standard. When the standard of reference is itself a stimulus with which members of the series are to be compared, the method is also called the *method of constant stimulus differences.*

It will be noted that these methods can be applied in fields extending beyond the sphere of simple perception. We shall not here discuss the necessary experimental control against errors due to differences in space and time and errors resulting from practice, fatigue, and mental attitude, and the statistical treatment of results obtained by the method of limits or the method of average error is so simple that it is adequately covered by the section on measures of central tendency, of dispersion, and of sampling error in the preceding chapter. We shall, however, review the mathematical principles in the "constant process" usually applied to results obtained by the method of right and wrong cases because it has a wider field of application, because it affords a more elaborate determination of the limen, and because its relation to the normal frequency distribution is not always clear to the student on account of a different terminology.

Suppose our problem is to determine the limen of dual impression on a certain part of the skin. Apparently we cannot expect to find an exact distance between two points which is at once necessary and sufficient for the observer invariably to perceive the points as two. The actual data obtained will be something like the following:

Distance in Paris lines:	0.5	1.0	1.5	2.0	3.0	4.0	5.0
% two-point judgments:	10	14	40	65	80	87	96

We are thus compelled to define the two-point limen as the magnitude of separation which yields 50% two-point judgments. It is seen to be somewhere between 1.5 and 2.0 Paris lines, at which there are, respectively, 40 and 65% two-point judgments. By linear interpolation between these two values, one obtains 1.70 as the limen. When the interval is not large, linear interpolation gives fairly satisfactory results, but the relation between the distance of separation and the proportion of two-point judgments is obviously not a linear function. If we designate the proportion of two-point judgments by p, the number of times in which a given stimulus is presented by N, the magnitude of the stimulus by D, the limen sought by L, and the standard deviation by σ, the following equation of an ogive will adequately represent the relation:

$$p = \frac{1}{N}\int_{-\infty}^{D-L} y \; dD \qquad [60]$$

It is generally assumed that the ogive is the integral of the normal distribution,

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{D-L}{\sigma}\right)^2} \qquad [61]$$

and if a perfect fit is obtained, the observed $(D—L)/\sigma$ value corresponding to each value of p will be equal to the theoretical x value as given in a table of the normal probability integral. The best fitting curve for any set of results will thus have such constants that the squares of the errors or discrepancies,

$$\left(x_1 - \frac{D_1}{\sigma} + \frac{L}{\sigma}\right)^2, \quad \left(x_2 - \frac{D_2}{\sigma} + \frac{L}{\sigma}\right)^2, \text{ etc.,}$$

will be a minimum. These quantities in summing must be weighted inversely as the squares of the standard errors of the x's, which are here percentile values. The standard error of a percentile as given by formula [14] is $i\sqrt{Npq}/f$. Since in a normal distribution $f = izN/\sigma$, it becomes $\sigma\sqrt{pq}/z\sqrt{N}$, where z is the ordinate. Accordingly, the weights are:

$$w_1 = \frac{N_1 z_1^2}{\sigma^2 p_1 q_1}, \quad w_2 = \frac{N_2 z_2^2}{\sigma^2 p_2 q_2}, \quad \text{etc.}$$

The σ is, of course, constant for the various stimuli and consequently may be disregarded without disturbing the relative weights. N depends upon the particular experiment and may also be dropped if the various stimuli are presented an equal number of times. Disregarding both N and σ, we have z^2/pq as the weighting factor in summing the squares of the errors. To determine the constants yielding the best fitting normal probability integral, then, we wish to make

$$\Sigma w \left(x - \frac{D}{\sigma} + \frac{L}{\sigma} \right)^2 = \Sigma \frac{z^2}{pq} \left(x - \frac{D}{\sigma} + \frac{L}{\sigma} \right)^2$$

a minimum. The condition is given by the normal equations,

$$\left.\begin{aligned} \Sigma wx - \frac{1}{\sigma} \Sigma wD + \frac{L}{\sigma} \Sigma w &= 0 \\ \Sigma wxD - \frac{1}{\sigma} \Sigma wD^2 + \frac{L}{\sigma} \Sigma wD &= 0 \end{aligned}\right\} \quad [62]$$

the solution of which yields:

$$\left.\begin{aligned} \frac{L}{\sigma} &= \frac{\Sigma wD \ \Sigma wxD - \Sigma wx \ \Sigma wD^2}{\Sigma w \ \Sigma wD^2 - (\Sigma wD)^2} \\ \frac{1}{\sigma} &= \frac{\Sigma w \ \Sigma wxD - \Sigma wD \ \Sigma wx}{\Sigma w \ \Sigma wD^2 - (\Sigma wD)^2} \end{aligned}\right\} \quad [63]$$

It is a general practice in psychophysical experiments to employ a measure of *precision, h,* instead of the standard deviation, σ. But the two are very simply related:

$$h = \frac{1}{\sigma\sqrt{2}} \qquad \sigma = \frac{1}{h\sqrt{2}} \quad [64]$$

The deviation values for various areas as given in Fechner's table are thus in terms of h and are equal to $x/\sigma\sqrt{2}$ if converted to values usually given in tables of the normal probability integral. It should be added that the weights given in Urban's table are in our terminology $\pi z^2/2pq$ values, the constant factor $\pi/2$ employed merely to make the maximum weight 1. Similarly, the weights given in Müller's table, which are incorrect and have been replaced by Urban's weights, are equal to $2\pi z^2$ and are simply the squares of the ordinates in terms of their maximum.

The statistical treatment of results obtained by the method of constant stimuli for the determination of a stimulus limen is without modification applicable to results obtained by the method of constant stimulus differences for the determination of a difference limen. The difference limen is either an upper or a lower limen, according as the difference, when added or subtracted from the magnitude of the standard, would make 50% of the cases yield "greater" or "less" judgments, respectively. The usual procedure permits the observer to give a third category of "equal" or "doubtful" judgments. Now the proportion of intermediate judgments which obviously is an important factor determining the limen depends upon the psychological attitude of the observer rather than his physiological sensitivity. Probably he can, if he is forced to, make "guesses" which will in the long run turn out to be more often right than wrong. If, however, the intermediate category of judgments be banned and the observer be forced to respond either "greater" or "less," the most probable magnitude of stimulus which will be judged half the time "greater" and half the time "less" than the standard will simply be equal to the standard, and both the upper and the lower limens under the old definition will vanish. It has therefore been suggested to make use of the interquartile range or simply retain the measure of precision, *h,* as determined from the dichotomy of judgments. Such measures will not be comparable to the limen, but will be better objective measures of sensitivity.

The normal probability integral is, of course, only one of the many possible functions which may be assumed to hold between the magnitude of stimuli and the proportion of positive judgments. Various formulae for non-linear interpolation may be used to obtain more accurate results than linear interpolation, and the $\tan^{-1}$ function has been tried. More experimentation is necessary before one can establish with any degree of assurance a mathematical function which is theoretically most reasonable and actually best fitting the observed facts. Thurstone has recently demonstrated that the relation between the proportion of judgments in one of two categories and the magnitude of the standard stimulus for comparison is a symmetrical function only if sensitivity is constant and independent of the stimulus magnitude, and that if sensitivity is inversely proportional to the magnitude of stimuli, as postulated by Weber's law, then the normal probability integral is unsound unless we employ in place of the stimulus magnitudes their logarithms.

Measures of Reliability and Validity

Suppose we devise a test to measure scholastic aptitude. The scores on the test (X) cannot be expected to represent the true magnitude of scholastic aptitude in purity (A), but will at least contain a certain amount of chance error (E_1). Thus

$$X = A + E_1$$

If they also contain a systematic error I we have the situation which is discussed in the next paragraph. The *coefficient of validity* of the test as a measure of scholastic aptitude is defined as the correlation between X

and A, and the *coefficient of reliability* of the test is defined as the correlation between X and a similar measure which has the same magnitude of chance errors. These two correlation coefficients depend entirely upon the relative magnitudes of the two factors, A and E_1:

$$r_{xa} = \frac{\sigma_a}{\sigma_x} \qquad [65]$$

$$r_{xX} = \frac{\sigma_a^2}{\sigma_x^2} \qquad [66]$$

But measures of A and E_1 are not directly available, and consequently the two correlation coefficients cannot be calculated. To overcome the difficulty, we have recourse to devising another test of scholastic aptitude, which contains chance errors as represented by E_2. Thus:

$$Y = A + E_2$$

If E_1 and E_2 are uncorrelated,

$$r_{xy} = \frac{\sigma_a^2}{\sigma_x \sigma_y} \qquad [67]$$

If, further, X and Y are comparable, i.e., E_1 and E_2 are of the same order of magnitude,

$$r_{xy} = \frac{\sigma_a^2}{\sigma_x^2} = \frac{\sigma_a^2}{\sigma_y^2} \qquad [68]$$

and it gives a measure of reliability for both of the tests. When we have only two tests, their comparability is not subject to verification, though more or less subject to control in their construction. But if we have more than two tests, we can derive more than one correlation coefficient, and their approximation to equality,

$$r_{12} = r_{13} = \ldots = r_{23} = \ldots$$

is a measure of the comparability of the different tests. When the intercorrelations vary considerably, it would involve an error to take for the reliability coefficient of a test its correlation with any of the others or even its average correlation with all the others. Under such conditions, the following formula gives a good estimate of the true reliability coefficient:

$$r_{xX} = \frac{(n-2)\,\bar{r}_{xp}^{\,2}}{n\bar{r}_{pq} - 2\bar{r}_{xp}} \qquad [69]$$

where the reliability coefficient of a test, r_{xX}, designates its estimated correlation with a comparable test, $\bar{r}_{xp}$ designates the average of the obtained correlations between the test and all the others, r_{pq} the average of all the obtained intercorrelations, and n the number of the tests.[5]

We have noted that equations [67] and [68] assume that E_1 and E_2 are uncorrelated. Under actual conditions, however, tests are likely to contain, in addition to chance errors which are uncorrelated, errors of a systematic order which are consistently present in the tests. Let us call the systematic errors I and the chance errors e. Thus:

$$X = A+I+e_1$$
$$Y = A+I+e_2$$

Then

$$r_{xy} = \frac{\sigma_a^2+\sigma_i^2}{\sigma_x\sigma_y} \qquad [70]$$

and the definition of the reliability of X becomes

$$r_{xX} = \frac{\sigma_a^2+\sigma_i^2}{\sigma_x^2} \qquad [71]$$

Here the relative magnitude of A and I are unknown. The correlation between X and Y attains unity as soon as $e_1 = e_2 = 0$, regardless of the presence or absence of A, not to say its magnitude. In other words, the two tests may have a perfect reliability in that they have perfect agreement, and yet neither is at all a measure of scholastic aptitude. So we need a criterion which could test whether X and Y are largely A or I. A perfect criterion of A is obviously unavailable. The best practicable procedure is to find a criterion which contains only errors independent of those of the tests. We may thus use records of academic achievement,

$$Z = A + e_3$$

or estimates by teachers or associates,

$$W = A + e_4$$

which are independent of the tests and will correlate with them only in so far as A is a common factor. Then

$$r_{xz} = \frac{\sigma_a^2}{\sigma_x\sigma_z} \qquad r_{xw} = \frac{\sigma_a^2}{\sigma_x\sigma_w}$$

Such correlations with imperfect criteria are often taken for the validity coefficient of a test. Obviously, this procedure involves a spurious concept, for it would make the validity of a test dependent upon the reliability of the criterion as well as upon its own excellence. Our definition of the validity coefficient, on the other hand, requires that we derive the correlation with a perfect criterion, which, though incapable of direct computation, is readily obtained by the formula,

---

[5] For the derivation of formula [69] and a discussion at greater length on the subject, see Shen (32).

$$r_{xa} = \sqrt{\frac{r_{xz}\ r_{xw}}{r_{zw}}} \qquad [65a]$$

which is reducible to equation [65], since

$$\sqrt{\frac{r_{xz}\ r_{xw}}{r_{zw}}} = \sqrt{\frac{\dfrac{\sigma_a^2}{\sigma_x\sigma_z}\ \dfrac{\sigma_a^2}{\sigma_x\sigma_w}}{\dfrac{\sigma_a^2}{\sigma_z\sigma_w}}} = \frac{\sigma_a}{\sigma_x}$$

With the use of the validity coefficient, r_{xa}, we can now estimate scores on A from scores on X by the regression equation,

$$\bar{a} = r_{xa}\frac{\sigma_a}{\sigma_x}x = \frac{\sigma_a^2}{\sigma_x}x = \frac{r_{xz}\ r_{xw}}{r_{zw}}x \qquad [72]$$

When the validity coefficient is not available, the best we can do is to estimate scores on $A+I$, by means of the reliability coefficient:

$$\overline{a+i} = r_{x(a+i)}\frac{\sigma_{a+i}}{\sigma_x}x = \frac{\sigma_a^2+\sigma_i^2}{\sigma_x^2}x = r_{xX}x \qquad [73]$$

The standard errors of estimate for the two regression equations are:

$$\sigma_{a\cdot x} = r_{xa}\sigma_x\sqrt{1-r_{xa}^2} \qquad [74]$$

$$\sigma_{(a+i)\cdot x} = \sigma_x\sqrt{r_{xX}-r_{xX}^2} \qquad [75]$$

Both of these standard errors, it will be noted, are smaller than the standard error of estimate as given by formula [24], which applies in estimating one series of actual measures from another. Thus, from the same obtained scores, we can estimate perfectly valid scores (which are free from any error) or perfectly reliable scores (which are free from chance errors) with a greater accuracy than imperfect scores (which contain both systematic and chance errors).

If we increase the samples of response in a test, that is, if we lengthen a test or combine it with other tests of the same thing, its reliability and validity are thereby increased. The combination of f comparable tests or f forms of the same test, has the following validity and reliability coefficients in terms of those of a single form:

$$r_{fx,\,FX} = \frac{fr_{xX}}{1 + (f-1)\,r_{xX}} \qquad [76]$$

$$r_{fx,\,a} = \frac{fr_{xa}}{\sqrt{f + (f^2-f)\,r_{xX}}} = \frac{r_{xa}}{\sqrt{\dfrac{1-r_{xX}}{f} + r_{xX}}} \qquad [77]$$

A special form of formula [76], where $f = 2$,

$$r_{2x,\,\mathrm{II}X} = \frac{2r_{xX}}{1 + r_{xX}} \qquad [76a]$$

is useful in estimating the reliability of a test, when no alternative form is available, from the correlation obtained between its two halves so divided that they are comparable to each other.

The presence of chance errors has the necessary effect of lowering the correlation obtained between two measures. To estimate the correlation which would obtain if the measures were perfectly reliable, we may follow the practice of Spearman and *correct for attenuation* by the formula,

$$r_{\infty_\omega} = \frac{r_{xy}}{\sqrt{r_{xX}}\sqrt{r_{yY}}} \qquad [78]$$

or by the more elaborate one,

$$r_{\infty_\omega} = \frac{(r_{xy}r_{xY}r_{Xy}r_{XY})^{\frac{1}{4}}}{\sqrt{r_{xX}}\sqrt{r_{yY}}} \qquad [78a]$$

If we use the Stanford Reading Test and the Thorndike-McCall Reading Scale for comparable measures of reading achievement, and the Stanford Arithmetic Test and the Buckingham Arithmetic Scale for comparable measures of arithmetic achievement, then the reliability coefficients, r_{xX} and r_{yY}, are the correlations obtained, respectively, between the two reading measures and between two arithmetic measures, the cross correlations, r_{xy}, r_{xY}, r_{Xy}, and r_{XY} are the observed "raw" correlations of each of the reading tests with each of the arithmetic tests, and the correlation corrected for attenuation, r_{∞_ω}, is the estimated correlation which would obtain between perfectly reliable measures of reading and arithmetic achievement, or between scores which are averages on an infinite number of comparable reading and arithmetic tests. Since formula [78*a*] makes use of all the four observed cross correlations, it may be expected to yield more reliable results than formula [78]. If the assumption of comparability holds, however, all the cross correlations should be equal and any one should be representative of all. Since the magnitude of the corrected coefficient is a ratio of the cross correlation to the reliability coefficient, it attains perfect unity when the observed cross correlations and the reliability coefficients have the same order of magnitude. Such a situation obtains when two things are so closely related that their measures are no more differentiable than different measures of the same thing. Occasionally one may obtain a corrected correlation coefficient which exceeds the limits of + 1.00, especially if the shorter formula [72] is used. Such paradoxical results are generally due to correlated errors present in the measures representing the two variables. If a Democrat and a Republican are both very much concerned with a certain issue in regard to which they have similar sentiments, they are likely to agree with each other better than the Democrats or the Republicans will among themselves.

Estimated correlation coefficients like r_{xa}, $r_{fx,a}$, $r_{fx,FX}$, and $r_{\infty\omega}$ have standard errors which differ from values given by formula [26]. They should be calculated by formulae especially derived for the purpose.[6] It should also be noted that since the correlation corrected for attenuation is what one would expect between perfectly reliable measures which are not actually available, there is obviously no occasion for its use in a regression equation for the estimation of one measure from another.

Sundry Problems Relating to Testing

From formula [77] it is seen that lengthening a test does not increase its validity unless the correlation between the different parts are imperfect. In other words, additional parts are altogether useless if they are entirely redundant. On the other hand, parts which are mutually prime in their contribution to the whole will have a validity much higher in combination than in isolation. It is therefore a general principle in test construction that the items selected for a test should correlate high with the criterion and low with each other.

The best weighting of parts of a test is given by the regression coefficients obtained with the technique of multiple correlation. Obviously, however,

---

[6]

$$\sigma_{r_{xa}} = \frac{r_{xa}}{2\sqrt{N}} \left[4r_{xa}^2 + \frac{2}{r_{xa}^2} + \frac{1}{r_{zw}^2} + \frac{1}{r_{xz}^2} + \frac{1}{r_{xw}^2} - \frac{2r_{xa}^2}{r_{xz}^2} - \frac{2r_{xa}^2}{r_{xw}^2} - 5 \right]^{\frac{1}{2}}$$

$$\sigma_{r_{fx,a}} = \left\{ \frac{r_{fx,a}}{r_{xa}} \sigma^2_{r_{xa}} + \frac{(1-f)\, r^3_{fx,a}}{2f r_{xa}^2} \sigma^2_{r_{xX}} \right\}^{\frac{1}{2}}$$

For the standard errors of $r_{\infty\omega}$ and $r_{fx,fX}$, see Kelley (13) and Shen (31).

it is inadvisable to apply multiple correlation to single items of which there are a very large number: aside from the enormous labor it would involve, the capitalization of chance errors would yield results which are anything but trustworthy, since the group upon whom a test is standardized is not likely to have a larger number of individuals than the number of items in a test. The general practice is to apply weighting only to groups of items which are manifestly dissimilar either in form or in content.

If the criterion of a test as well as the response to a question is forced into a dichotomy in scoring, the ϕ as defined in formula [36] is used which is the product-moment correlation coefficient of a four-fold table. The formula in the present case may be written in the following form:

$$\phi = \frac{n_{1a}\ n_{2b} — n_{1b}\ n_{2a}}{\sqrt{n_1\ n_2\ n_a\ n_b}} \qquad [36a]$$

where n_1 and n_2 designate frequencies of the columns, n_a and n_b frequencies of the rows, and n_{1a}, n_{2b}, n_{1b}, and n_{2a} frequencies of the cells. In a test of professional attitudes, Cowdery and Strong weighting the various questions in proportion to $\phi\ /\ (1—\phi^2)\sigma$ have obtained very good results in differentiating between the legal, medical, engineering, and other professions. Here the correlation with the criterion is represented by ϕ, its trustworthiness by $1/(1—\phi^2)$, and the standard deviation of the item as scored by σ.

In scoring a test, one may either credit a person with all items attempted, or only with items correctly done, or penalize a person for incorrect responses. Letting x designate the number of items attempted and y the number of items incorrectly done, we can write the score as

$$x — cy$$

where an optimum value for c is to be determined. A priori, c should be the ratio of guesses to incorrect responses. Assuming that a person tested takes all the opportunities to guess, we see that, if the questions are of the true-false type in which he makes a choice between two possible answers, he would probably have made two guesses when giving one wrong answer. Similarly, if the correct responce is a choice of one from three, one would have made three guesses for every two wrong answers. In general, if the correct answer is a choice of one from n possible responses, the ratio of guesses to wrong answers is given by $n/\ (n—1)$, which is the value for c. It will be seen that c has an upper limit of 2 when n is 2, and a lower limit of 1 when n is infinite. The latter situation is given by questions of the recall type in which no possible answers are suggested and therefore only incorrect responses may be attributed to guessing.

In case the foregoing analysis relating to guessing cannot be assumed for a test situation, one can resort to an empirical method by which c is determined such as to yield the highest coefficient of validity or reliability (17, pp. 93-96). It is hardly necessary to point out that if a test is given with sufficient time to allow each individual of a group to complete the questions, x becomes constant and the value of c becomes immaterial since changes in its magnitude will not change the relative standing of the different individuals.

Both the difficulty of a problem and the achievement of an individual can always be stated in terms of percentiles, i.e., respectively, by the proportion of individuals who fail to answer the question correctly or who fail to attain the proficiency of the given individual. For purposes of comparison and correlation, it is often required to convert such percentile scores into deviation scores. If normality of distribution can be assumed, the necessary conversion is readily accomplished by referring to a table of the normal probability integral. The resulting scores will be positive or negative according as they are above or below the median of the distribution, and theoretically they have no limit in magnitude since the normal curve does not intersect the base line within finite distances.

The above method with modification may be applied in converting ranks to scores on a linear scale. For utmost accuracy of results, each rank should be regarded as a portion of the area under a normal curve and its mean deviation from the mean determined by:

$$d = \frac{z_1 - z_2}{q_1 - q_2} \tag{79}$$

where q_1 and q_2 are the areas beyond the lower and upper limits of the portion in question, and z_1 and z_2 are the corresponding ordinates in a unit normal distribution of the two boundaries. Thus the score for the third rank in a total of 50, which is the score for the area between the 94th and 96th percentiles, is given by the difference between the ordinates at these percentiles divided by .02, which is the mean deviation of the area under consideration. If one uses the deviation of the median, instead of the mean, of the area (in the present case the 95th percentile for the third of 50 ranks), the difference is insignificant except for the two extremes and except when the number of ranks is small. Reference to a table of the normal probability integral yields 1.647 for the mean deviation of the area between the 94th and 96th percentiles[7] and 1.645 for the median deviation of the same area or the 95th percentile.

Those who abhor the use of negative and fractional quantities may, in place of a normal curve with its mean at 0 and its σ equal to 1, follow McCall (21) and use a normal curve with a mean at 50 and a σ of 10, or use Hull's curve (9), which has a mean at 50 and a σ of 20, and which is normal except that it is truncated at 2.5σ above and below the mean.

Since frequencies in a normal distribution are concentrated near the mean and scattered at the extremes, the conversion of ranks or percentiles into a linear scale based on such a distribution will show that a difference between the 5th and 10th percentiles is several times greater than one between the 45th and 50th percentiles. Of course, uncertainty as to the normality of distribution affects to a like degree the trustworthiness of the converted scale. But direct comparison or correlation between ranks or percentiles is not a sound procedure unless the ranks or percentiles are assumed to be equidistant or rectangularly distributed.

---

[7] $(.119123 - .086174)/.02 = 1.647$.

The use of tests for simple measurements of aptitude or proficiency should present no serious problem. The individual score is considered in the light of its error of estimate as given by formulae [24], [74], and [75] and in relation to the central tendency and dispersion of the group in which it is found. In comparing measures in two distributions, however, we need to know the standard errors of certain differences which have not yet been given.

Suppose we desire to compare the stature of men with that of women. A distribution of differences could be obtained from a comparison of every man with every woman, and the mean and standard deviation of the distribution could be determined. But these are readily derived from the means and standard deviations of the two separate distributions for men and women. The mean of the differences is simply equal to the difference of the means, while the standard deviation of the differences is calculated to be as follows:

$$\sigma_{(x-y)} = \sqrt{\sigma_x^2 + \sigma_y^2 - 2r_{xy}\sigma_x\sigma_y} \qquad [80]$$

Since there is no correspondence between the individual measures in the two distributions, r_{xy} is 0, and the last term under the radical can be discarded. But if we are considering the difference between brothers and sisters, or between husbands and wives, we expect a positive correlation which is due to heredity or selection and which may substantially reduce the standard deviation of the differences. Since the standard error of the mean is given by the standard deviation of the distribution divided by $\sqrt{N}$, formula [80] divided by $\sqrt{N}$ gives the standard error of the mean difference. The same result can be obtained by considering x and y as the means and σ_x and σ_y as their standard errors; then $\sigma_{(x-y)}$ would become the standard error of the difference of the means. The formula then becomes

$$\sigma_{M_x-M_y} = \sqrt{\frac{\sigma_x^2}{N} + \frac{\sigma_y^2}{N} - 2r_{M_xM_y}\frac{\sigma_x\sigma_y}{N}} \qquad [80a]$$

wherein for ordinary situations $r_{M_xM_y}$ is known not to vary systematically from r_{xy} so that r_{xy} may be used in its place.

Another problem of comparison or differences for which we need a standard-error formula is indicated by such questions as "How much more has John learned in history than in science?" or "How much better is Robert likely to succeed as a scientist than as a politician?" Obviously the proficiency or aptitude of the given individual must be considered in relation to other human beings. We have in the preceding chapter already referred to the use of standard measures,

$$z = \frac{\sigma}{x} = \frac{X - M}{\sigma}$$

which give the positions of individuals in a distribution independent of the unit with which the measurement happens to have been made. The difference between John's achievements in history and science, or the difference between Robert's aptitudes for science and politics, is then measured by the difference between such z scores, and the significance of the difference is seen when it is stated in terms of its standard error,

$$\sigma_{(z_1 - z_2)} = \sqrt{2 - r_{1I} - r_{2II}} \qquad [81]$$

which is a function of the two reliability coefficients.

The maze which has been very extensively and successfully employed in animal experiments to determine the part played by various receptors and the effect due to different surgical lesions often appears to be a rather doubtful instrument for measuring individual differences. The reliability of maze scores is difficult to evaluate because we do not have the same assurance here as in the case of most human tests either in regard to comparability between different patterns or in regard to absence of correlated errors between different sections of a maze or between different trials on it. Most patterns seem to allow chance factors to enter in such a complex fashion that very little of anything stable is tested. However, recent investigations with the multiple-T pattern have yielded rather high correlations, and the maze gives greater promise than ever before of being a fruitful experimental apparatus (37).

Differentiable Abilities

Practically all the tests that are available at present are empirical tests constructed and applied without a comprehensive analysis of the abilities of man. Even tests of "general intelligence" are nothing but juxtapositions of heterogeneous elements combined to give a correlation with age, school grade, or estimates of teachers. While this procedure is obviously quite fruitful for immediate specific results, it is hardly the most economical in the long run. And it cannot be expected to yield analytical results which will add to our knowledge of human nature. From the latter point of view, by far the most significant problem is to determine the various dimensions or independent factors in which human beings should be measured. Once these are known, and their means of measurement are available, their combination to give specific scores, such as on scientific aptitude or poetic appreciation or travelling salesmanship, will be a relatively simple matter.

Thorndike has advanced the suggestion that human abilities should be measured in three respects: height or level of difficulty, extent or variety of different tasks, and speed. Level and speed are quantitative concepts easily amenable to measurement. But as soon as we consider extent, we are confronted with the endless varieties of tasks and with the consequent problem of determining whether any two are equivalent or distinct or mere variations of each other. To settle this question we have recourse to correlation. Each task can be geometrically represented by an axis: perfectly correlated tasks by the same axis and uncorrelated tasks by axes at right angles. The level and speed of a task are then simply magnitudes along an axis; if they are not perfectly correlated, the two scores must be considered results of two different tasks and be represented on two different axes. Thus we have a system of n dimensions; any individual can be located by n independent measures from which can be derived various scores which are combinations of any two or more of the n measures. A complete system will undoubtedly have a very large number of dimensions which we cannot hope to determine in a few years. But we are reasonably certain that only a relatively few of these will figure with any prominence in all ordinary lines of human activity.

According to the *theory of two factors* in its pure form and not as Spearman and others have modified it to fit certain relevant facts, there is a general factor g which participates in all correlated forms of intelligent behavior. Furthermore, every sort of ability tested can be divided into two factors: a general factor g which is the same throughout all sorts of ability and a specific factor s which is uncorrelated with g and uncorrelated with any other s. In other words, all intercorrelations found to exist between different tests are attributable to g and to g alone. This proposition is contingent upon the condition

$$r_{12}r_{34} = r_{13}r_{24} = r_{14}r_{23} \qquad [82]$$

i.e., upon the condition that all *tetrad differences*—differences between any two of the three products in the equation—are equal to zero. Equation [82] can be derived from the assumption that the partial correlation between any two tests with g constant,

$$r_{xy\cdot g} = \frac{r_{xy} - r_{xg}r_{yg}}{\sqrt{1 - r_{xg}^2}\sqrt{1 - r_{yg}^2}}$$

is zero, thereby yielding

$$r_{xy} = r_{xg}\ r_{yg}$$

so that each of the three products in equation [82] is equal to $r_{1g}\ r_{2g}\ r_{3g}\ r_{4g}$. Spearman has at various times used less rigorous criteria than the equality of tetrad differences to zero, about which there has been considerable controversy. The use of the tetrad difference requires its evaluation in terms of its sampling error for which formulae derived have not gained unanimous assent. Refraining from lengthy mathematical arguments in the present discussion, we may mention several lines of general criticism.

In the first place, there is no warrant that the g found in different tables of correlation coefficients is constant. On the contrary, there are reasons to believe that *g* has several components which enter in various degrees in the different correlation tables investigated. Here we come to certain fundamental conditions necessary for the analytic study of abilities; namely, a large number of variables must be employed in a single investigation, and the sample studied must be large and homogeneous.

In the next place, while *g* is a sufficient explanation for situations which satisfy the criterion of tetrad differences, it is by no means a necessary one. An alternative system is the *theory of sampling,* suggested by Thomson (1), according to which the variables are factored into numerous independent elements which are distributed in a chance fashion. The two-factor theory and the sampling theory are thus compatible to results derived from the same correlations; conflict will occur only in their interpretation of the psychological and physiological mechanism.

Spearman qualifies his theory by limiting the *g* to participate only in all branches of "intellectual" activity and by making an exception to the complete mutual independence of specific factors in case two or more operations are "similar." He has explicitly recognized the existence of several group factors. In the light of these qualifications, many of the destructive criticisms against the two-factor theory seem to have been quite unnecessary. At the same time it becomes evident that the designation of a factor as "general," "group," or "specific" has no significance except in the particular situation under investigation.

There are several factors besides *g* which Spearman recognizes as having a general character. One of these is a general mental *inertia* or *perseveration*, the negative aspect of which Garnett has called *"c"* or *cleverness,* which is closely associated with "quickness" and "originality." Another appertains to the *oscillations* of mental energy, probably manifestations of fatigue. Still another factor is the factor *"w"* in the field of conation. It has variously been called the *persistence of motives, purpose,* or *self-control.* Spearman interprets these factors by an analogy of energy, engines, and the engineer. The factor *g* is a measure of the quantity of intellectual energy possibly having a correlate in the cerebral cortex. Both inertia and oscillation can then be regarded as further aspects of the same energy, while the various specific factors are engines in which the energy operates, and, finally, the conative factor *w* has the rôle of the engineer. In addition to these, Spearman has found certain group factors necessary to account for correlations in excess of what can be attributed to *g*. These include a "logical ability," a "conjunction," or arithmetic ability, a "constructive mechanical ability," "music appreciation," and varieties of "memory."

One of the writers using an extension of Spearman's technique in a series of tests has found a general factor which is attributable to *heterogeneity* of *maturity, sex,* and *race.* Five independent group factors were also established in more than a single population: these factors have to do with *verbal* material, *number* or *quantitative* concepts, *memory, spatial* concepts, and *speed*, respectively. Two additional factors were found in some of the populations to which a larger number of tests were applied:

one of these may be termed *ebullience* or *vivacity*, and the other is a second spatial factor which has to do with the *manipulation of spatial relations* as distinguished from their apprehension and retention.

As any of an indefinite number of coordinate systems may serve for a frame of reference, we cannot upon mathematical principles alone establish any one set of factors as the proper set. Common sense, however, would definitely prefer a set of factors which are fewest in number, most reliable in measurement, and capable of simplest descriptions. The analytic determination of independent factors in various intercorrelated abilities is a very promising field of mathematical as well as experimental research, and studies of such traits will lead to conclusions of psychological interest and of biological and social importance.

REFERENCES

1. Brown, W., & Thomson, G. H. Essentials of mental measurement. London: Cambridge Univ. Press, 1921.
2. Cowdery, K. M. Measurement of professional attitude differences between lawyers, physicians, and engineers. *J. Person. Res.*, 1926, **5**, 131-141.
3. Elderton, W. P. Frequency curves and correlation. New York: Layton, 1927.
4. Fisher, R. A. The distribution of the partial correlation coefficient. *Metron*, 1924, **3**, 329-334.
5. ———. Statistical methods for research workers. London: Oliver & Boyd, 1925.
6. Forsyth, C. H. An introduction to the mathematical analysis of statistics. New York: Wiley, 1924.
7. Garrett, H. E. Statistics in education and psychology. New York: Longmans, Green, 1926.
8. Holzinger, K. J. Statistical methods for students in education. Boston: Ginn, 1928.
9. Hull, C. L. Aptitude testing. Yonkers: World Book, 1928.
10. Kelley, T. L. Tables to facilitate the calculation of partial coefficients of correlation and regression equations. *Bull. Univ. Texas*, 1916, No. 27.
11. ———. Chart to facilitate the calculation of partial coefficients of correlation and regression equations. Stanford Univ.: Stanford Univ. Press, 1921.
12. ———. A new measure of dispersion. *J. Amer. Statis. Asso.*, June, 1921.
13. ———. Statistical method. New York: Macmillan, 1928.
14. ———. A new method for determining the significance of differences in intelligence and achievement scores. *J. Educ. Psychol.*, 1928, **14**, 321-333.
15. ———. The influence of nurture upon native differences. New York: Macmillan, 1928.
16. ———. The interpretation of educational measurements. Yonkers: World Book, 1927.
17. ———. Crossroads in the mind of man: a study of differentiable mental abilities. Stanford Univ.: Stanford Univ. Press, 1928.
18. Kelley, T. L., & McNemar, Q. Doolittle versus the Kelley-Salisbury iteration method for computing multiple regression coefficients. *J. Amer. Statis. Asso.*, 1929, **24**, No. 166.
19. Kelley, T. L., & Salisbury, F. S. An iteration method for determining multiple correlation coefficients. *J. Amer. Statis. Asso.*, 1926, **21**, 282-292.

20. KEYNES, J. M. A treatise on probability. London: Macmillan, 1921.
21. MCCALL, W. A. How to measure in education. New York: Macmillan, 1922.
22. MILLS, F. C. Statistical methods applied to economics and business. New York: Holt, 1924.
23. OTIS, A. S. Statistical method in educational measurement. Yonkers: World Book, 1925.
24. PEARL, R. Introduction to medical biometry and statistics. Philadelphia: Saunders, 1927.
25. PEARSON, K. [Ed.] Tables for statisticians and biometricians. London: Cambridge Univ. Press, 1924.
26. PINTNER, R. Intelligence testing: methods and results. New York: Holt, 1928.
27. RIETZ, H. L. Mathematical statistics. Chicago: Open Court Publ. Co., 1927.
28. ——————. [Ed.] Handbook of mathematical statistics. Boston: Houghton Mifflin, 1924.
29. RUGG, H. O. Statistical methods applied to education. Boston: Houghton Mifflin, 1917.
30. SALISBURY, F. S. A simplified method of computing multiple correlation constants. *J. Educ. Psychol.*, 1929, **20**, 44-52.
31. SHEN, E. The standard error of certain estimated coefficients of correlation. *J. Educ. Psychol.*, 1924, **15**, 462-465.
32. ——————. The reliability coefficient of personal ratings. *J. Educ. Psychol.*, 1925, **16**, 232-236.
33. SHIROKOGOROFF, S. M. Process of physical growth among the Chinese. Shanghai: Commercial Press, 1925.
34. SPEARMAN, C. The abilities of man: their nature and measurement. New York: Macmillan, 1927.
35. SPEARMAN, C., & HOLZINGER, K. The sampling error in the theory of two factors. *Brit. J. Psychol.*, 1924, **15**, 17-19.
36. ——————. Note on the sampling error of tetrad differences. *Brit. J. Psychol.*, 1925, **16**, 86-88.
37. STONE, C. P., & NYSWANDER, D. B. The reliability of rat learning scores from the multiple-T maze as determined by four different methods. *Ped. Sem.*, 1927, **34**, 497-524.
38. SYMONDS, P. M. Measurement in secondary education. New York: Macmillan, 1927.
39. THOMSON, G. H. A direct deduction of the constant process used in the method of right and wrong cases. *Psychol. Rev.*, 1919, **26**, 454-464.
40. ——————. A new point of view in the interpretation of threshold measurements in psychophysics. *Psychol. Rev.*, 1920, **27**, 300-307.
41. ——————. The tetrad-difference criterion. *Brit. J. Psychol.*, 1927, **17**, 235-255.
42. THORNDIKE, E. L. Mental and social measurements. New York: Teach. Coll. Bur. Publs., 1913.
43. THORNDIKE, E. L., *et al.* The measurement of intelligence. New York: Teach. Coll. Bur. Publs., 1926.
44. THURSTONE, L. L. The learning curve equation. *Psychol. Monog.*, 1919, **26**, No. 3.
45. ——————. The fundamentals of statistics. New York: Macmillan, 1925.
46. ——————. The phi gamma hypothesis. *J. Exper. Psychol.*, 1928, **11**, 293-305.

47. Titchener, E. B. Experimental psychology. Vol. II. Quantitative experiments. New York: Macmillan, 1905.

48. Tolley, H. R. & Ezekiel, M. The Doolittle method for solving multiple correlation equations versus the Kelley-Salisbury iteration method. *J. Amer. Statis. Asso.*, 1927, **22**, 497-500.

49. Webb, E. Character and intelligence. *Brit. J. Psychol., Monog. Suppl.*, 1915, No. 3.

50. Whittaker, E. T., & Robinson, G. The calculus of observations. New York: Van Nostrand, 1924.

51. Yule, G. U. An introduction to the theory of statistics. London: Griffin, 1927.

NAME INDEX

Abbott, 361, 387
Abelsdorff, 240, 241, 271
Abney, 188, 266, 269
Abraham, 293
Abramowski, 515
Achmatov, 277, 307, 310
Adami, 36
Adams, 559
Adrian, 136, 137, 139-142, 152, 155, 159, 164, 225, 226, 228, 252, 253, 261, 269, 350, 351, 387, 394, 397-399, 405, 411
Aggozzotti, 538
Alexander, 764
Allen, 188, 212, 293, 309, 312, 326, 331, 349
Allison, 362, 363, 388
Allport, F. H., 739, 766-768, 784, 807
Allport, G. W., 768
Alrutz, 404, 411
Alston, 404, 411
Alvarez, 461-463, 484
Alverdes, 746, 784
Ames, 179, 212
Ammon, 36
Anderson, 675, 780
Andova, 118, 122
Andrade, 175, 212
Anrep, 423
Appunn, 277
Arey, 50, 53, 55, 60, 78, 122-124, 352, 372, 388
Aristotle, 407, 449, 490, 498, 686
Arnstein, 357, 388
Arps, 681
Arrhenius, 120, 121, 227, 228, 233
Arthur, 667, 701
Asai, 359, 388
Assheton, 36
Aubert, 246, 269
Aveling, 515, 516
Ayres, 711, 712, 718, 721, 723, 735, 764
Azara, 29

Babinski, 828, 831
Backman, 363, 388
Bagg, 6, 39
Baginsky, 323, 347
Bailey, 477, 484, 485
Baillarger, 543
Bainbridge, 460, 484
Baird, 199, 212
Baldi, 50, 123
Baldus, 50, 123
Baldwin, B. T., 646, 648, 659
Baldwin, J. M., 740, 779
Ball, 549, 563
Ballard, 608-610, 622
Bancroft, 123
Banister, 273-349
Barany, 424
Barcroft, 468
Bard, 449-487, 521
Barnes, 114, 115
Barr, 830
Barral, 382, 388
Bartlett, 281, 310-312, 791, 807
Barton, 347
Basedow, 650
Bateson, 36
Batson, 571, 573, 575, 622
Bazett, 472, 473, 484
Bechterew, see Bekhterev
Becker, 388
Beer, 388
Beers, 830
Bekhterev, 474, 484
Bell, 492
Bentley, 532
Bentley, M., 408, 411
Benussi, 506, 521
Bere, 701
Bergson, 499, 525, 544
Beritoff, 538, 562
Bernard, 460
Bernstein, F., 36
Bernstein, J., 138, 164
Bertholon, 514
Best, 234, 269
Bethe, 388
Bezold, 197, 199, 201
Bezold, von, 264
Bichat, 460
Bickel, 461, 467, 484
Bidder, 445, 447
Bills, 200, 212
Binet, 616, 648, 659, 661-663, 667, 669, 676, 683, 686, 690, 695-697, 699-701, 707, 717, 735, 758
Bingham, 784
Binswander, 515
Bird, 529, 562, 566, 623
Bishop, G. H., 137, 139, 164, 394, 395, 412
Bishop, H. C., 293, 311
Blaauw, 220, 269
Blackford, 752
Blakeman, 852
Blanchard, 246, 269
Blanton, 659
Blatz, 503, 505, 506, 521
Bleuler, 830
Bliss, 457, 485
Blix, 401, 411
Bloch, 220, 269
Blondel, 221, 270
Boas, 661
Bode, 279, 311
Boldyreff, 117
Boll, 240, 270
Bolton, 530, 537, 661
Boltzman, 281
Book, 573, 574, 606, 607, 623, 701
Borgquist, 500, 521
Boring, 319, 322, 324, 339, 340, 347, 402, 404, 411
Boscovitch, 59
Botezat, 411
Bouman, 548, 562
Bourdon, 209, 212, 607, 623
Boyd, 648
Boysen-Jensen, 123
Bramwell, 139, 164
Brauer, 36
Braunstein, 220, 270
Bray, 611, 614, 623
Breed, 529, 566, 625
Brewer, 36
Brewster, 96
Bridges, 40
Brigham, 701
Bright, 457, 485
Brill, 831
Brissaud, 476
Britton, 453, 457, 464-467, 469, 471, 479, 481, 484, 485
Broca, 200, 212
Brodhun, 246, 249, 271
Broman, 360, 388
Bronk, 159, 164, 269

Bronner, 669, 685, 701, 702, 735, 763, 784
Brown, E. M., 625
Brown, T. G. 253, 270, 561
Brown, W., 582, 596, 610, 623
Brown, Wm., 881
Brown-Séquard, 3, 4, 36,
Brücke, 143, 159, 165, 197, 199, 201
Brünn, von, 356, 357, 360, 388
Brunswick, 141, 156, 165, 504, 509, 521
Bryan, 572-575, 623
Burnett, C. H., 315, 317, 348
Burnett, M. C., 404, 412
Burnside, 646, 659
Burt, C., 685, 686, 701, 784
Burt, C. L., 735
Burtt, 506, 521, 681, 682, 701
Butler, 2, 37
Buyse, 701
Buytendijk, 532
Buzby, 494, 521

Cajal, 401, 412, 541, 556
Calkins, 565, 623
Camis, 154, 164
Campbell, 412
Campos, 468, 484
Cannon, 117, 434-448, 453-459, 461-468, 470-472, 475, 479-485, 501, 504, 508, 510, 513, 521
Capper, 645
Carlson, 117
Carlson, A. J., 438, 439, 447
Carmichael, 527, 562, 566, 623
Carr, 402, 412, 501, 521, 587, 588, 620, 623
Carr-Saunders, 5, 38
Cason, 551, 558, 562, 780
Cassamajor, 2
Castle, 123
Castlemajor, 529
Cattell, H., 150, 165
Cattell, J. McK., 661, 707, 759, 765, 784
Cattell, McK., 142, 150, 165
Chadwick, 709, 735
Champy, 37
Chapman, 616, 619, 623, 735, 784
Charpetier, 186
Charters, 712, 735
Child, 164, 527, 528, 562, 743, 744, 769-772, 781, 782
Church, 830
Claparède, 122
Clark, 58, 123
Clarke, 759
Cleeton, 752
Clements, 578, 580, 623
Cobb, P. W., 175, 184, 212
Cobb, S., 150, 159-161, 165, 167, 471
Cobbey, 408, 413
Coblentz, 183, 213
Coghill, 635, 659
Cohn, 382, 388
Cole, 49, 56, 57, 78, 123, 124
Colvin, 686
Comstock, 174, 213
Congdon, 2, 39
Conklin, 830
Cooley, 740, 784
Cooper, 142, 154, 165
Coover, 613, 623
Cope, 37
Cornetz, 115
Corning, 735
Corti, 287, 306, 317, 323, 332
Cosens, 328, 348
Cotton, 830
Courtis, 709, 735
Cowdery, 875, 881
Cowdry, 388
Cox, 759, 784
Crandell, 303, 311
Crew, 37
Crile, 557
Crozier, 48-127, 239, 270, 350-391, 533
Cuénot, 37
Cummins, 587, 623
Cunningham, 673-675, 684
Cunningham, J. T., 5, 37
Cushing, 370, 371, 388
Cushny, 475, 485
Cutolo, 403, 412

Dallenbach, 402, 404, 412, 603-605, 624
d'Allonnes, 518, 521
Dana, 481-483, 485
Danforth, 37
Daniels, 121
Darrow, 516, 568-570 603, 625
Darwin, 3, 5, 460, 490-493, 497, 501, 521, 698, 748
Dashiell, 784
Davenport, 96
Davies, 402, 413
Davis, H., 128, 141, 142, 156, 165, 168
Davis, H. B., 607, 611, 623
Day, 532
Day, I. S., 703
Dearborn, 667, 675, 701
Decroly, 701
Deenik, 282
Deiters, 426
deKleyn, 646
Delage, 5, 37
de la Paz, 465
Denisoff, 648
Denny-Brown, 154, 165
Denworth, 701
Descartes, 449, 561
Decoeudres, 648, 659
Detlefsen, 2, 37
Detwiler, 527, 528, 562
Deutschmann, 37
DeVoss, 736
Dewey, 807
DeWick, 514, 522
Dickson, 701
Dieterici, 214, 266, 271
Dimmick, 172, 173, 210, 213, 403, 412
Dingfelder, 37
Dittmers, 184, 213
Dobrowolsky, 264, 270
Dodge, 87, 213, 430-433
Dolley, 124
Donaldson, 401, 412
Doolittle, 859
Dorcas, 379, 390
Downey, 646
Downing, 143, 165
Dreschel, 126
Driesch, 544
Drozynski, 506, 521
duBois-Raymond, 514
Duchenne, 492, 493
Dugdale, 698, 701
Dumas, 462, 485, 493, 498, 499, 505, 521
Duncker, 94, 96

Dunlap, 429, 432, 549
Dupuy, 37
Dürkin, 6
Duser de Barenne, 470, 471, 485

Ebbecke, 558, 562
Ebbinghaus, 557, 561, 565, 566, 584, 585, 600, 603, 605, 606, 608, 623, 661
Eberhardt, 290, 311
Eddington, 629
Edelmann, 323
Ehrlich, 357, 388
Eimer, 37
Eisentraut, 391
Elderton, 881
Eldridge, 2, 37
Elliot, 713, 737
Elliott, 453, 466, 485
Ellis, H., 759, 831
Ellis, R. S., 784
Emerson, 183
Emery, 37
Eng, 502, 521
Ephrussi, 582, 623
Erffmeyer, 682, 701
Erlanger, 137, 139, 164-166, 394, 395, 412, 504
Estabrook, 698, 702
Eustachius, 450
Everett, 316, 348
Evringham, 153, 167
Ewald, J. R., 318, 324, 325, 348
Ewald, W. F., 125
Ewart, 38
Ewert, 400, 406, 412, 611, 612, 623
Exner, 266, 270, 557
Ezekiel, 859, 883

Fairchild, 38
Farabee, 38
Fechner, 55, 72, 183-186, 245, 252, 270, 417, 868
Federighi, 55, 60, 61, 78, 86, 106, 118, 124
Feleky, 494, 505, 521
Fenn, 143, 165
Fenton, 735
Féré, 514
Fernald, 708, 736
Fernandez, 29, 38
Ferree, 200, 213
Field, 143, 159, 165
Figurin, 648
Findlay, 5, 38
Fischer, C. R., 708, 737
Fischer, E., 28, 363, 388
Fisher, G., 708, 709
Fisher, R. A., 859, 881
Fletcher, 278, 280, 283, 286-288, 292, 303-305, 311
Flügel, 307-311
Foa, 515
Forbes, 128-168, 253, 270, 392, 412
Forel, 38
Forsyth, 881
Foster, 648, 660, 663, 704
Foucault, 536
Fourier, 274, 285
Fowler, 277, 280, 281, 311
Fracker, 613, 623
Frantz, 38
Franz, 400, 406, 412, 546, 562, 809-831
Franzen, 735
Frappa, 493, 521
Fraser, 322, 348
Frazer, 792, 807
Freeman, 690, 696, 702, 705-737, 765, 784
Freud, 500, 501, 739, 817, 828
Frey, 316, 348
Frey, von, 363, 370, 389, 401, 402, 404, 412
Freyd, 784
Fries, 67, 125
Frisch, von, 360, 361, 384, 389, 534
Fröhlich, 366, 389
Fryer, 760, 784
Fuchs, 404, 412
Fulton, 145, 159, 166, 469, 477, 484, 485
Fusfeld, 703

Galen, 449
Galton, 26, 34, 38, 662, 697, 698, 702, 747, 748, 758, 784, 816, 831
Garman, 359
Garnett, 880, 881
Garrey, 49, 125
Garten, 240, 270
Garth, 682, 702
Gaskell, 451, 452, 455, 485
Gasser, 137, 166, 394, 395, 412
Gates, 649
Gates, A. I., 597, 598, 623, 731, 736
Gates, G. S., 494, 497, 522
Gates, R. R., 38
Gelb, 546, 562
Gellhorn, 363, 389
Gerard, 143, 144, 153, 165, 166
Gerbach, 648, 659
Gertz, 206, 213
Gesell, 628-660, 669, 702
Giard, 38
Giersberg, 389
Gilbert, 661
Gildemeister, 516
Gilliland, 736
Gley, 535
Goddard, 698, 763, 830
Goldberg, 221, 270
Goldscheider, 401, 402, 404, 412, 558, 559
Goldsmith, 534
Goldstein, 544, 546, 562, 563
Golgi, 368
Goltz, 470, 471, 474, 486, 558
Goodenough, 648, 659, 667, 702
Gordon, H., 696, 702, 765
Gordon, K., 596, 623
Gotch, 141, 166
Gowen, 38
Gowin, 750, 751, 784
Graff, 507, 522
Grassé, 119
Graves, 695, 702
Gray, 325, 343, 348, 349, 465, 736
Greene, 736
Gregg, 150, 159, 165, 270
Gregory, 736
Griffith, C. R., 2, 38, 432
Griffith, F. R., Jr., 156, 165
Gross, 831
Grünbaum, 548, 562
Guernsey, 347, 348
Gueroult, 290
Gullette, 503, 522
Guyer, 5, 38

Hachet-Souplet, 554
Haggerty, 675, 699, 702
Haggqvist, 402, 412
Haig, 234
Halverson, 659

Hammarberg, 530, 537
Hancock, 278, 311
Hänig, 389
Hanström, 360, 389
Harden, 602, 623
Hardesty, 317, 348
Harrison, 5
Harter, 572-575, 623
Hartline, 125
Hartman, 467
Hartridge, 313-349
Hartshorne, 763, 784
Hartwell, 360, 389
Harvey, 636
Hayes, 685, 702
Hays, 127
Head, 402, 412, 472, 482, 486, 514, 517, 522, 547, 548, 563
Healy, 685, 702, 708, 736, 763, 784, 831
Hecht, 55, 57, 61, 87, 101, 103, 105, 106, 125, 169, 174, 180, 181, 188, 204, 205, 213, 216-272, 372, 386
Heidenhain, 368, 389
Heidrich, 370, 389
Heilman, 702
Heine, 603, 623
Held, 334
Hellwig, 556
Helm, 208
Helmholtz, 179, 213, 263, 264, 266, 271, 273, 282, 288, 290, 293, 295, 309, 311, 316, 317, 324, 329, 331, 332, 334, 335, 341, 348
Hempelmann, 122
Henmon, 536, 686, 736
Henning, 465, 366, 372, 389
Henri, 661
Hering, 525
Hering, E., 1, 38, 103, 174 201, 213, 244, 264, 266, 271, 306
Hering, J., 29
Herlitzka, 373, 389
Heron, 568, 625
Herrick, 367, 368, 389, 544, 563
Herring, 663, 702
Herter, 122
Hertwig, G., 38
Hertwig, O., 6, 38
Herzog, 388
Hess, 185, 213, 239, 271
Hetzer, 645, 648
Heusen, van, 372, 391
Hill, 763
Hill, A. V., 136, 143, 145, 146, 162, 165, 166, 619, 623
Hines, 736
Hinsey, 159, 166
Hippel, von, 5
Hirsch, 702
Hitchcock, 122
Hoagland, 61, 78, 118, 125
Hobbes, 498, 499, 740, 779
Hobhouse, 575, 623
Hoch, 821, 831
Hochstetter, 5
Hoeber, 389
Hoelzel, 440, 447
Hoke, 737
Hollingworth, H. L., 389, 773, 784
Hollingworth, L. S., 702, 736, 755
Holmes, G., 546, 563
Holmes, S. J., 49, 119, 122, 125
Holsopple, 414-433
Holt, 211, 213
Holzinger, 881
Hooker, 527
Hooker, H. D., 239, 272
Hopkins, A. E., 354, 389
Hopkins, A. McH., 141, 156, 165
Hopkins, L. T., 702
Horn, 711, 736
Houssay, 474, 486
Howard, 445, 447
Howell, 158, 166, 355, 368, 371
Hubbert, 588, 590, 624
Hughlings Jackson, 472, 547, 548
Huguenin, 610, 624
Huijsman, 308
Hull, 876, 881
Hunter, T. A., 530, 563
Hunter, W. S., 433, 564-627, 649
Huntington, 26
Hurst, 319, 348
Huxley, 5, 38
Hyatt, 39

Irwin, 702
Ives, 185, 187, 195, 200, 201 213, 266, 271
Ivy, 118
Izquierdo, 456, 468, 486

Jackson, 472, 547, 548
Jacobsen, 538
Jacobsen, C., 503, 522
Jacobson, 358, 360
Jaensch, E. R., 173, 213
Jaensch, W., 650
Jagodowski, 357, 358, 389
James, 478-484, 486, 491, 492, 508, 509, 518, 822
Jamieson, 683, 702
Janet, 831
Jastrow, 661
Jefimoff, 387
Jelliffe, 830
Jenkins, J. G., 603-605, 624
Jenkins, R. L., 39
Jenkinson, 39
Jennings, 66, 122, 125, 129, 166, 524, 532, 535, 563, 578, 624
Jensen, Boysen-, 123
Jhering, von, 39
Johannsen, 27, 39
Johnson, B. J., 646, 667, 702
Johnson, H. M., 558, 559, 563, 648, 659
Jolly, 151, 166
Jones, 281, 282, 285, 305, 311
Jones, D. F., 39
Jones, H. E., 646, 692
Jones, L. A., 189, 213
Jones, M. C., 645, 646 649, 659
Jones, N. F., 711
Jordan, R. H., 736
Jorgensen, 736
Jost, 585, 624
Judd, C. H., 614, 624, 785
Judd, D. B., 266, 271
Jukes, 698
Jung, 515, 830
Just, 125

Kafka, 122
Kahlenberg, 370, 373, 374, 389
Kallikak, 698
Kammerer, 5, 39
Kant, 498, 499
Kantor, 739, 785

Kappers, 527, 556-558, 560, 563
Karbe, 402, 413
Karplus, 474, 486
Karpman, 361, 391
Karwoski, 199, 201
Kato, 141, 142, 167
Katz, D., 173, 213
Katz, S. H., 362, 363, 388
Katzaroff, 597, 598, 624
Keeler, 73, 125
Keesing, 805, 808
Keith, 332-335, 348
Kellaway, 466
Kelley, 702, 733, 736, 832-883
Kempf, 529
Kent, 667, 703
Keynes, 882
Kidd, 39
Kiesow, 372, 389, 402, 404, 412
Kincaid, 615-618, 624
King, 440
Kingsbury, 675
Kingsbury, B. A., 288, 289, 311
Kingsbury, E. F., 213
Kirilzev, 477, 486
Kirkpatrick, 702
Kirschmann, 185, 200, 213
Kitson, 785
Kjerstad, 568, 569, 591, 624
Klüver, 550, 563
Knight, 752, 784
Knors, 597, 624
Knudsen, 277-279, 281, 282, 285, 305, 311, 339, 348, 349
Kodama, 466, 486
Koehler (see Köhler)
Koenig, 240, 246, 249, 252-254, 257-260, 263, 266, 271 (see also König)
Koettgen, 240, 241, 271
Koffka, 544, 554, 563, 577, 624, 649, 659
Köhler, 383, 575-578, 580, 624, 649
Kohlrausch, 230, 235, 271, 279, 342
Kohs, 649, 667, 702
Kolmer, 334
König, 203, 214 (see also Koenig)
Köttgen (see Koettgen)
Koyanagi, 5
Kraepelin, 39, 814, 821, 831
Kranz, 305, 312
Krasnogorski, 649
Krause, 404
Kreidl, 474, 486
Kremer, 363, 389
Kretschmer, 751, 785
Kribs, 372
Kries, von, 255, 257, 258, 263, 264, 271
Kropp, 80, 90, 125
Krueger, 605, 624
Kucharski, 293, 311
Kuehne, 240, 271
Kuhlmann, 648, 659, 663, 669, 675, 702
Kühn, 70, 122, 597, 624
Kühne (see Kuehne)
Kuile, 338
Kundt, 240, 244, 271
Kussmaul, 644, 645

Lacy, 682, 703
Ladd, 207
Ladd-Franklin, 172, 206, 214, 264, 266, 271
Lakenan, 582, 624
Lamarck, 1
Lamson, 468
Landis, 488-523
Lane, 277, 311
Lange, 478-481, 483, 484, 486, 502, 509, 517
Langfeld, 494, 522
Langford, 188, 201, 215
Langley, 267, 271, 451, 452, 455, 459, 460, 486
Lapicque, 136, 158, 167, 513, 514, 522, 537, 563
Larguier des Bancels, 363, 389, 582, 624
Larson, 504, 522
Lasareff, 220, 235, 253, 271, 384, 386, 387, 389
Lashley, 94, 524-563, 587, 588, 624
Laurens, 239, 272, 432
Lavater, 490
Le Dantec, 39
Lee, A., 41
Lee, F. S., 153, 167
Lehmann, 404, 412
Lendl, 39
Lenhossék, von, 368
Lennox, 160, 161, 167
Lenz, 39
Leonard, 650
Leslie, 39
Levin, 143, 167
Lewis, 453, 457, 464, 479, 485
Leyton, 545
Libby, 55, 124
Liddell, 150, 151, 167
Liebermann, A., 360, 390
Liebermann, P. V., 325, 339, 348
Lillie, 39, 138, 139, 152, 167, 390
Lincke, 210, 214
Lincoln, 736
Lippman, 646
Lister, 546
Little, 6, 39
Liu, 588, 590, 624
Locke, 817
Loeb, 48, 50, 53, 57, 117, 122, 125, 126, 147, 167, 230, 239, 272, 558, 563
Löhner, 119
Lotze, 400, 412
Lovén, 370, 390
Lowry, 189, 213
Lucas, 136, 137, 140, 147, 152, 155, 156-158, 161, 164, 167, 253, 272, 355, 390
Luciani, 335, 348
Ludgate, 752, 785
Luft, 278, 311
Luh, 610, 624
Lundin, 468, 485
Lush, 39
Lutz, 153, 167
Lyer, 206
Lyon, 595, 596, 624

MacAuliffe, 39
MacCurdy, 160-162, 167, 831
MacDowell, 4, 39
Mach, 174, 214, 314, 348
Maciesza, 39
Mackenzie, 288, 289, 311
MacLeod, 452
MacPhail, 680, 703
Mader, 292, 311
Magnus, 156, 161, 167, 470, 486, 646
Malinowski, 808
Maloney, 380
Mangelsdorf, 124
Mangold, 118
Mann, 505, 522

Mantegazza, 495
Marañon, 480, 511, 522
Marbe, 210, 214
Mark, 310, 311
Marks, 702
Marston, L. R., 649, 659
Marston, W. M., 503, 522
Martin, C. L., 438, 447
Martin, E. G., 144, 167
Martin, M. F., 173, 214
Mast, 50, 125, 126, 239, 272, 531
Mateer, 649, 659
Matthaei, 563
Matthes, 390
Matthews, 225, 269, 557
Maxwell, 266
Maxwell, J. C., 46, 126
Maxwell, S. S., 122, 432
May, M. A., 738-785
May, R. M., 390
Mayer, A., 446, 448
Mayer, A. M., 293, 311, 330, 331, 348
Mayer, L., 522
McCall, 675, 677, 703, 725, 736, 873, 876, 882
McCarthy, 544
McCordock, 2, 39
McCouch, 412
McDonald, 558
McDougall, 2, 3, 39, 186, 214, 559, 740, 768, 779, 785, 830
McGeoch, 607, 608, 624
McGinnis, 645
McIndoo, 360, 361, 390
McIver, 457, 485
McKendrick, 314, 324, 334
McKenzie, 495
McNemar, 859, 881
Meek, 659
Meenes, 408, 413
Melton, 607, 608, 624
Mendel, 3, 8-10, 14-16, 18, 20, 22-24, 28, 36, 39
Mendelsohn, 535
Mendenhall, 465
Metalnikow, 532
Meumann, 555, 624
Meyer, 559
Meyer, A., 831
Meyer, M. F., 311, 319, 333, 348, 400, 412
Meyer, S., 39
Michaelis, 122
Miles, A. L., 432
Miles, M., 39
Milinelli, 474
Miller, 675
Miller, D. C., 284, 285, 311
Miller, E., 735
Miller, R. H., 150, 165
Mills, 882
Minkowski, 644
Minnich, 126, 360, 361, 384, 390
Minton, 348
Mitchell, 62, 126
Mohr, 6, 40
Monakow, von, 546, 563
Monroe, 736
Moore, 48, 59, 67, 124, 126
Morax, 211, 214
Morgan, C. L., 162, 167
Morgan, J. J. B., 830
Morgan, L. H., 800, 808
Morgan, T. H., 1-44, 47, 126
Morgulis, 48, 59, 126
Moseley, 566, 624
Moss, 759
Mosso, 502, 503
Mott, 820
Müller, 206, 308, 868
Müller, A., 103, 126
Müller, G. E., 565, 587, 600, 625
Müller, H., 126
Muller, H. J., 6, 35, 40
Müller, J., 266, 267, 272, 397, 399, 400, 412
Müller, L. R., 439, 441, 442, 448
Munk, 546
Münsterberg, 661, 708, 736
Murchison, 763, 764, 785
Murray, E., 405, 413
Murray, G. R., 40
Myers, 382, 383, 390, 675
Myers, C. S., 290, 299, 300, 312
Myerson, 40, 518, 522

Naccarati, 751
Nafe, 393-413, 517, 522
Nagel, 230, 272, 366, 390
Nash, 699, 702
Navez, 53, 67, 79, 124, 126
Nernst, 136, 167, 241
Nesfield, 304, 312
Neumann, 582, 625
Newman, 29, 30, 40
Newton, 263
Nice, 648
Nicolai, 604, 610, 625
Nikiforowsky, 292, 312
Norsworthy, 595, 625, 661
Northrop, 57, 126
Nutting, 184, 214
Nyswander, 878, 882

Obreshkove, 230, 272
Odell, 736, 737
Odin, 759
Oehrwall, 370, 372, 390
Oertly, 382, 383, 390
Ogburn, 798, 808
Ogden, 741, 783, 785
Oglivie, 40
Olmsted, 159, 165, 353, 372, 390
Oppenheimer, 126
Orbeli, 159, 167, 469
Ornstein, 40
Orr, F. J., 42
Orr, H. B., 40
Orth, 40
Orton, 648
Osborn, 40
Osterhout, 354, 379, 390
Ostwald, 138, 167
Otis, 675, 882
Oxnard, 55, 73, 124

Paal, 126
Pack, 445, 448
Packard, 6, 41
Papanicolaou, 43
Paris, 867
Parker, 41, 48, 113, 122, 126, 129-133, 135, 143, 168, 350-391
Parsons, 197, 202, 214
Pascual, 165
Paterson, 588, 625, 666-668, 684, 703, 752, 785
Patten, 126
Patterson, 29, 30, 40
Pattie, 308, 310, 312
Paul, 375, 378, 391
Paulsen, 361
Paulu, 703, 737
Pauly, 5, 41
Pavlov, 4, 324, 445, 448, 508, 535, 552, 563, 566, 625
Pearce, 612, 625
Pearl, 882

Pearson, 27, 36, 41, 96, 427, 698, 703, 850, 852, 860, 861, 882
Pechstein, 535, 563, 581, 583, 625
Peckham, 745
Peiper, 646, 660
Pendleton, 413
Penfield, 472, 473, 484
Pentschew, 582, 625
Penzoldt, 363, 388
Perry, 446, 448
Perserico, 515
Peterson, 515, 645
Peterson, G. M., 572, 625
Peterson, J., 400, 406, 413, 554, 563, 571, 621, 625, 662, 682, 686, 703
Piaget, 648, 659
Pick, 547
Pictet, 41
Piderit, 494
Piéron, 118, 185, 186, 214, 221, 272, 515, 563
Pike, 432
Pilzecker, 565, 587, 600, 625
Pincus, 61, 63, 73, 74, 86, 90, 91, 97, 99, 124, 126
Pincusson, 126
Pintner, 550, 661-704, 762, 763, 785, 882
Piper, 230, 233, 234, 272
Pitt-Rivers, 803, 806, 808
Plato, 686
Poffenberger, 389, 614, 615, 625
Pohlman, 305, 311
Polimanti, 119
Politzer, 305, 314, 315, 348
Popenoe, 41
Poppelreuter, 525, 546, 548, 554, 556, 563
Porteus, 667, 703
Poulton, 41
Powell, 467
Pratt, 140, 168
Pressey, L. C., 675, 703, 737, 830
Pressey, S. L., 675, 703, 737, 830
Preston, 175, 214
Pretori, 185, 213
Preyer, 644, 660
Priest, 176, 192, 195, 214
Proctor, 179, 212
Przibram, 41
Purdy, 198, 199
Purkinje, 186, 205, 214
Pusch, 531
Pütter, 220, 235, 272
Pyle, 560, 582, 595, 625

Quain, 176, 177, 214, 368
Querido, 156, 158, 168, 457, 481, 485

Rabaud, 118, 119
Radecki, 515
Radossawljewitsch, 605, 606, 625
Rainey, 645
Ranc, 382, 388
Rand, 200, 213
Ranson, 133, 402, 413
Ranvier, 368
Rapport, 466, 473, 485
Rasmussen, 802, 808
Ray, 156, 165
Rayleigh, 281, 293, 299, 312, 348
Raymond, 41
Raymond, duBois-, 514
Read, 357, 360, 391
Reed, 581, 582, 625
Reeves, 179, 214, 221, 249, 257, 272
Reh, 41
Rehwoldt, 506
Reid, 42
Reidmayr, 42
Reissner, 316, 319
Renqvist, 391
Renshaw, 400, 406, 413, 504
Retzius, 357, 368, 370, 391
Reusch, 391
Révész, 277, 312, 348
Rey, 221, 270
Rice, J. M., 709, 710, 728, 737
Rice, L. H., 142, 143, 165, 166, 168, 412
Richards, 374, 391
Richter, 118, 126
Riesz, 282, 312, 348
Rietz, 882
Rignano, 42, 525
Rivers, G. H. L. F. P. (See Pitt-Rivers)
Rivers, W. H. R., 208, 214, 412, 832
Roback, 675
Robertson, 557
Robinson, 122
Robinson, E. S., 568-570, 600-603, 625
Robinson, G., 883
Rochat, 264, 272
Roelofs, 203, 214
Rogers, 538
Rogers, F. T., 438, 447
Rogoff, 466, 467
Romanes, 3, 42
Rosanoff, 42, 821, 828, 830, 831
Roscoe, 218, 220, 270
Rose, 126
Rosenthal, 42
Ross, 808
Rothmann, 470, 471, 486
Roughton, 333
Roux, 8, 42
Rubaschkin, 356, 357, 391
Ruch, 587, 625, 703, 737
Ruckmick, 494
Ruedin, 42
Rugg, 721, 737, 882
Russell, 184, 214
Rutenburg, 220, 272
Rutherford, 318, 348
Ryder, 42

Sacher, 304, 312
Saito, 466, 487
Salisbury, 859, 881, 882
Sandiford, 683, 699, 702
Sasaki, 461, 467, 484
Sataké, 466, 486
Schaeffer, 355, 391
Schäfer, 291, 312
Schäfer, E. A., 323, 324, 348
Schafer, E. S., 457, 464, 486
Schiff, 460, 486
Schilf, 515
Schjelderup, 58, 202, 214
Schjelderup-Ebbe, 747
Schmidt, 445, 447
Schopenhauer, 498
Schröder, 42
Schuberth, 515, 516
Schultze, 357, 358
Schuyler, 573, 626
Schwalbe, 370, 391, 426
Scott, 745
Scripture, 611, 625
Sears, 575, 626
Seashore, 498, 522, 708 735, 737, 757
Secérov, 42
Seguin, 830

Semon, 2, 5, 42, 525
Severins, 119
Sewall, 308, 309, 312
Shakow, 667, 703
Shambaugh, 323, 349
Sheldon, 366, 391
Shen, 832-883
Shepard, 566, 625
Sheppard, 841, 860
Sherman, I., 645
Sherman, M., 496, 497, 501, 522, 645
Sherrington, 145, 148-154, 156-158, 161, 165, 167, 168, 204, 215, 460, 472, 473, 479, 480, 486, 487, 501, 524, 545, 557, 561, 772
Shinn, 630
Shirokogoroff, 862-864, 882
Shohl, 468
Siemens, 42
Silorsky, 505
Simon, 648, 659, 663, 735
Simpson, 734
Sims, 762
Sinden, 192, 215
Sioli, 42
Skaggs, 506, 597, 601, 625
Slawson, 763, 785
Sleight, 614, 626
Slight, 505, 506, 522
Smith, E. A., 5, 38
Smith, E. M., 281, 312
Smith, F. O., 192, 215
Smith, H. L., 737
Smith, T. L., 625
Snyder, J. C., 582, 625
Snyder, L. H., 42
Sollier, 43
Sommer, 43
Soury, 460, 486
Southard, 814, 821
Sowton, 156, 168
Spearman, 170, 215, 427, 687, 688, 700, 703, 707, 733, 734, 737, 850, 873, 879, 880, 882
Spencer, 1, 43, 499, 769, 778
Sprowls, 808
Stabler, 354, 391
Starch, 585, 586, 611, 626, 713, 737
Stark, 126
Starr, 504, 523
Stecher, 643, 659
Stedman, 703
Steffens, 581, 582, 626
Steinach, 529
Stenquist, 708, 737
Stepanoff, 323, 349
Stern, 309, 648, 660, 686, 703
Stewart, G. N., 466, 467, 486
Stewart, G. W., 279, 299, 312
Stier, 58, 61, 67, 70, 98, 115, 124, 126
Stiles, 132, 153, 168
Stockard, 6, 43
Stoddard, 616, 626
Stone, C. P. 529, 563, 588, 591, 592, 594, 626, 878, 882
Stone, C. W., 709, 737
Störring, 830
Stratton, 400, 406, 413
Stricht, Van der (see Van der Stricht)
Strieck, 391
Strong, 875
Strughold, 402, 413
Strümpel, 404, 413
Stucker, 278, 312
Stumpf, 278, 295, 312
Sturtevant, 40
Stutsman, 648, 660, 669, 703
Sugawara, 466, 486, 487
Sullivan, A. H., 408, 413
Sullivan, W. C., 831
Sulzer, 200, 212
Swift, E. J., 571, 573, 607, 611, 626
Symonds, 703, 882
Szily, von, 5, 43

Tachanoff, 514
Talbot, 187, 201
Taliaferro, 126
Tarde, 740, 779
Tashiro, 143, 168
Taussig, 759, 760
Taylor, N. W., 374, 375, 377, 379, 380, 391
Taylor, W. S., 830
Terman,, 663, 674, 675, 679, 686, 692, 699, 703, 732, 737, 765, 785
Thomas, D. S., 785
Thomas, J. A., 43
Thomas, W. I., 785
Thompson, 534
Thompson, H., 659
Thompson, S. P., 122, 293, 314, 349
Thomson, 880, 881, 882
Thorndike, 551, 563, 575, 576, 578, 579, 592-594, 611, 613, 620-622, 626, 674, 675, 680-682, 686, 687, 689-692, 700, 703, 704, 709, 712, 721, 723, 737, 758, 763, 785, 873, 878, 882
Thorsen, 550, 563
Thucydides, 686
Thurstone, 619, 626, 631, 660, 675, 723, 737, 869, 882
Tilney, 529
Titchener, 170, 215, 396, 400, 403, 405, 413, 490, 517, 523, 883
Toepler, 281
Tolley, 859, 883
Tolman, 4
Toops, 675
Torrey, 127
Totten, 507, 523
Toynbee, 314, 349
Trabue, 675, 737
Travis, L. E., 281, 312, 530, 563
Travis, R. C., 433
Tredgold, 830
Trendelenburg, 240, 241, 264, 265, 272
Trevesz, 538
Troland, 169-215, 263, 266, 272
Trotter, 402, 413
Trow, 575, 626
Tudor-Hart, 645
Turner, 119, 534, 563
Tylor, 792, 793, 808

Uexhüll, von, 288
Uhthoff, 258, 272
Ulrich, 585, 586, 626
Uno, 511, 523
Upton, 95
Urban, 868
Urbantschitsch, 293, 307, 312

Valentine, 293, 312
Vance, 278, 312
Van der Stricht, 357, 391

Van Wagenen, 737
Ven, 533
Veraguth, 515
Veress, 363, 391
Verlaine, 391
Verschaeffelt, 96
Verworn, 150, 168
Vesalius, 449, 450
Vicari, 4, 39, 43
Vigouroux, 514
Vincent, 567-570, 591, 613, 626
Viola, 751
Voeste, 201, 215
von Frey (see Frey, von)
von Kries (see Kries, von)
von Lenhossék (see Lenhossék, von)

Wacker, 360, 391
Wada, 117
Waite, 467
Wald, 234
Walduyer, 334
Walker, 468, 484
Waller, 318, 349, 515
Wallin, 830
Walther, 122
Ward, 2, 43
Warden, 585, 587, 626
Warren, 766, 785
Washburn, A. L., 117, 438, 447
Washburn, M. F., 549
Washburn, R. W., 649
Wasteneys, 272
Watanabé, 466, 486, 487,
Watson, 32, 43, 122, 497, 501, 523, 578, 579, 588, 589, 605, 620, 626, 645, 649, 660, 738, 766, 767, 785
Watt, 301, 312, 319, 335-339, 349
Weaver, 196
Webb, E., 883
Webb, L. W., 613, 626
Weber, 55, 245, 252, 272, 278, 282, 363, 391, 407, 413, 417, 503, 869
Wechsler, 515, 758
Wedensky, 154-156, 158, 161, 168
Wedgewood, 698
Wegel, 277, 280, 281, 305, 311, 312
Weigert, 244, 272
Weinberg, 293, 309, 312, 326, 331, 349
Weismann, 4, 8, 43, 744
Weiss, 127, 558, 563
Wells, 704, 830
Wertheimer, 209, 210, 215, 544, 554
Westphal, 189, 215
Wettstein, von, 43
Whatley, 682, 702
Wheeler, 43, 45, 127, 384, 391, 746, 776, 785
Whipple, 626, 704
Whitaker, 150, 165
White, 502
White, W. A., 830
Whitely, 601, 626
Whittaker, 883
Wien, 282
Wiggam, 43, 785
Wiggers, 459, 487
Wiggins, 784
Wilder, 44
Wiley, 506, 522
Wilkinson, 324-326, 333, 349
Williams, H. B., 412
Williams, O., 610, 627
Williams, R. E., 181, 213, 240, 241, 271
Willis, 450
Wilson, 829
Wilson, G. M., 712, 737
Wilson, H. A., 299, 300, 312
Wilson, J. G., 433
Wilson, S. A. K., 475-477, 481, 487
Wiltbank, 613, 627
Winch, 581-583, 596, 627
Windle, 44
Wissler, 707, 737, 786, 802
Witasek, 597, 627
Wolf, 648
Wolf, E., 57, 87, 90, 99, 103, 106, 124, 127, 259, 271
Wood, 681, 704, 737
Woodrow, 361, 391, 704
Woods, 44
Woodworth, 207, 472, 473, 487, 550, 594, 611, 613, 626, 627, 740, 766, 773, 780, 785
Woody, 737
Woolley, 708, 737
Wright, 468
Wright, W. W., 737
Wrightson, 317-319, 348, 349
Wrzosek, 39
Wunderli, 404, 413
Wundt, 371, 391, 490-492, 517, 661
Wylie, 613, 627

Yagi, 57, 127
Yarbrough, 612, 624
Yealland, 831
Yerkes, 533, 536, 563, 577, 578, 580, 627, 648, 660, 663, 692, 704
Yoshii, 323, 324, 349
Young, K., 785, 808
Young, P. T., 296, 302, 312
Young, T., 263, 264, 266, 267, 272, 291, 335
Yule, 27, 854, 859, 883

Zeleny, 44
Ziegler, 557
Ziegler, E., 44
Ziegler, H. E., 44
Zigler, 408, 413
Zoth, 205, 215
Zotterman, 142, 143, 164, 166, 252, 269, 394, 397, 398, 411
Zsymanski, 119
Zwaardemaker, 361, 362, 364, 366, 391

SUBJECT INDEX

Abilities, differentiable, 878
  distribution, 730
  general, 661
    individual differences, 750
    vs. special, 733
  initial and final, 614
  special, constitution, 730
    vs. general, 705
    measured by tests, 732
    measurement, 705
  types, 733
  (See also *Intelligence*)
Abnormal psychology, 809
Abnormality, four types, 810
  matter of degree, 810
Accommodation, visual, 178
Achievement *vs.* special aptitude, 706
Achromatic colors, 171
Acid taste, stimulus, 374
Acquired characters, inheritance, 1
Acuity of skin, 408
  visual, and illumination, 253
      stimuli characteristics, 204
Adaptation, chromatic, 201
  dark, 101, 230
  in feeling sense, 408
  light, 101, 229
  olfactory, 363
  time, 393
  visual, 187
Adjustor organ, definition, 130
Adult psychology, differential, 747
Affect, definition 489
Affective qualities in felt experience, 410
After-discharge, 151
After-sensations, auditory, 293
Age, calibration, 631
  experimental significance, 629
  and learning, 588
  and mental diseases, 816
  norms, 726
  prenatal periods, 632
  scores, 724
  of stimulus method, learning, 565
Agnosia, visual, and cerebral function, 546
Alcoholism and mental disease, 818
Algebraic summation in reflex, 156
Alienation, coefficient, 848
Alignment Chart, 859
All-or-none principle, 140, 252
    in heredity, 9
    statement, 142
Aments, 810
Anaesthesia and emotion, 475
  susceptibility of reflex, 154
Analysis without interference in animal behavior, 61
Anatomy of ear, 313, 421
    eye, 177
    facial expression, 492
  individual differences, 748
  of labyrinth, 421
  race differences, 753
  variations, social significance, 750
Anhedonia, 518
Animal experiments and hearing theory, 323
  psychology, differential, 744
  societies, individual differences, 745
    integration, 774
Animals, higher, integration, 771
  lower, differentiation, 742
    integration, 769
Antagonism of parasympathetic and sympathetic, 459
Anthropology, methods, 788
  and social psychology, 788
Aphasia and cerebral function, 546
  semantic, 548
  syntactical, 548
Appetite *vs.* hunger and thirst, 434
Appropriation, sensory, 353
Apraxia and cerebral function, 546
Aptitude, special, 706
  tests, origin and development, 706
  (See also *Ability*)
Arches of Corti, 317
Arithmetic mean, 837
Arrhenius equation, 227
Association of analogous sensation, principle, 491
Attainment, initial and final, 614
Attenuation, correction, 873
Audition, individual differences, social significance, 757
  maximum intensity for, 281
  minimum intensity for, 281
  volume, 301
  (See also *Hearing*)
Auditory after-sensations, 293
  fatigue, 305
    absolute threshold, 307
  localization, 295
  space perception, 295
  and tactual sensations, 280
Autokinetic sensation, 211
Autonomic nervous system, antagonism of two divisions, 459

control of viscera, 460
divisions, 451
in emotion, 513
functions, 455
general description, 134
Average deviation, 840

Bar diagram, 836
Basilar membrane, anatomy, 317
Beats, 289
  binaural, 292
Behavior, adaptive, and learning in infants, 648
  basis, 128
  emotional, 449
  infant, language, 646
    observation at Yale Psycho-Clinic, 638
    range, 642
  mathematical formulation, 82
  modifiability, 87
  patterns, biogenetic, 634
    classification, 633
    clinical study, 636
    comparative study, 637
    correlated, 634
    durational, 634
    experimental analysis, 635
    of genetic sequence, 634
    in infancy, 633
    integration, 160
    investigation, 635
    maturational basis, 650
    of maturity, 634
    method of co-twin control, 637
    naturalistic observation, 636
    neurophysiology, 635
    origin during growth, 525
    prehension, and maturation, 651
    psychometry, 635
    specific, 634
  perceptual, in infant, 645
  and physics, 45
  physiological foundations, 770
  prenatal, studies, 644
  social, primitive, 791
  and temperature, 55
  tropistic, compounding, 67
  variability, 87
Bezold-Brücke phenomenon, 197
Binaural beats, 292
Binet scales, 663
Binocular color mixture, 204, 264
  criteria of depth, 206
  flicker, 204
Biological background of differentiation, 742
Bitter taste, stimuli, 379
Blind, tests, 685
Blood changes in emotion, 504
  pressure in emotion, 503
  sugar in emotion, 468, 504
  supply, dependence of reflex, 154
  types, heredity, 24
  volume in emotion, 502
Blue arc phenomenon, 205
Blushing, 493
Bodily activity in emotion, 501
Bone conduction in audition, 291
Boring's hypothesis of hearing, 339
Borrowing customs and practices, 795
    procedure, 797
Brain, 133
  primitive parts involved in emotion, 470
  spot hypotheses of mental disease, 814
Bright-adaptation, 187
Brightness (See *Brilliance*)
Brilliance, 171
  absolute threshold, 183
  contrast, 185
  function of intensity, 183
    spatial factors, 185
    time, 186
      wave-length, 180
      composition, 185
  growth and decay, 186
  psychophysics, 179
  and size, 185
Bulky color, 173

c factor in intelligence, 880
calcarine cortex in vision, 211
Capacities, general, individual differences, 750
Catatonic dementia precox, 823
CAVD test, 687
Central mechanisms for emotional reactions, 469
  nervous system, 134
    functions, 146
  processes in habit, 549
  tendency, measures, 837
Cerebral cortex in vision, 211
    in emotion, 513
Change, culture, 798
Charpentier's bands, 186
Chemical changes at cell contact and learning, 557
  senses, 350
Chemoreceptors, 353
  common characteristics, 354
Children, handicapped, tests, 684
  superior, relatives, 698
  (See also *Infant*)
Choc, definition, 488

Chroma, function of intensity, 197
    time, 200
    wave-length, 191
  scale, 190
  spatial determinants, 198
  (See also *Color*)
Chromatic adaptation, 201
  attributes, psychophysics, 189
  colors, 171
  contrast, 200
  flicker, 200
  zones, 199
Chromosomes, 7
  conjugation and reduction, 10
  X and Y, 29
Chronaxie, 136, 393, 513
Circus movements, 55
Citizenship and intelligence, 763
Classification of odors, 365
Cleverness a general factor in intelligence, 380
Clinical crib, 638
Coaching in intelligence tests, 694
Cochlea, anatomy, 316
Coefficient of alienation, 848
    contingency, 854
    correlation, product-moment, 847
    kurtosis, 860
    regression, 847
    reliability, 870
    skewness, 860
    validity, 869
    visibility, 181
College student, intelligence tests, 680
Color, binocular relations, 203
  blind, visual acuity, 258
  blindness, 202
  bulky, 173
  chromatic and achromatic, 171
  complementary, 191
  film, 173
  function of wave-length, 197
  mixture, binocular, 204, 264
    three-component, 192
    triangle, 194, 195
    two-component, 191
  properties, 171
  solid, psychological, 171
  surface, 173
  theory, Hecht, 263
    Hering, 264
    Ladd-Franklin, 264
    Young, 263
  zones, 199
  (See also *Chroma*)
Combination tones, 289, 290
Common chemical sense, 350, 366
Communication, 780
Community, adult in, 738
  as culture unit, 790
  human, integration, 777
  and social life, 790
Complementary colors, 191
Completeness of response, learning, 621
Conditioning of infants, limitations, 657
Conducting substance and learning, 557
Conduction, 135
  bone, in audition, 291
  with decrement, 141
  membrane theory, 138
  properties, 513
  time, 151
Conflict of cultures, 799
    inter-group, 805
  and survival of cultures, 786
Connection of movement with sense ideas, principle, 491
Consonance, 289, 294
  theories, 295
Contact of culture, 795
    individual reactions, 804
Contiguity, law, 551
Contingency, coefficient, 854
  mean square, 853
  method, 852
  square, 853
Contractile function, development, 129
Contraction, muscular, 144
Contrast, brilliance, 185
  chromatic, 200
Correction for attenuation, 873
    guessing in scoring, 726
Correlation, contingency method, 852
  curvilinear, 850
  multiple, 857
  partial, 855
    tables, 859
  product-moment, 847
  ratio, 850
  rectilinear, 846
  of variation in nerve discharge and felt experience, 394
Cortical retina, 548
Co-twin control, method, 637
Cretinism, heredity, 24
Criteria of emotion, 489
Crossed extension reflex, 149
Crossing over in heredity, 20
Crying and laughter, 498
Culture, adjustment of subject group, 802
  change, 798
  conflicts, inter-group, 805
  determinants, 791
Cultures, conflict, 799
    and survival, 786
  contact, 795

    individual reactions, 804
  reaction, 794
  survival, 806
Current of injury, 137
Curve-fitting, 859
Curve, growth, 836
  learning, 567, 836
  leptokurtic, 861
  memory, 609
  mesokurtic, 861
  normal probability, 836
  ogive, 836
  platykurtic, 861
Curvilinear correlation, 850
Cutaneous receptors, 392
  sensitivity, 392
Cyclothymia, 826

D, 840
Damping, 329
Dark-adaptation, 187, 230, 234
Darwin's principles of emotion, 490
Deaf, tests, 684
Deafness, 303, 304
Death feigning, 118
Decay and growth of brilliance, 186
Decorticate animals and emotion, 471
Decrement, conduction with, 141
Defects, optical, 179
Delayed instincts, 529
Delerium tremens, 818
Delusions, forms, 821
Dementia praecox, 821
  catatonic, 823
  hebephrenic, 822
  paranoid, 824
  and thyroid, 814
Density in nerve impulse discharge, 394
Dependents, intelligence, 762
Depth, monocular and binocular criteria, 206
Determinants of culture, 791
Deuteranopia, 202
Developmental zero, 631
Deviation, average, 840
  mean, 840
  quartile, 840
  standard, 841
Diagnosis, educational tests, 729
Dichromatism, 202
Diencephalic process theory of emotion, 483
Diencephalon as seat of emotion, 473
Differences, individual, in animal societies, 745
  race, in tests, 682
Differentiable abilities, 878
Differential adult psychology, 747
  animal psychology, 744
  effects of environment, 743
  social psychology, 742
Differentiation, biological basis, 742
Digestion and emotion, 460
Dimming effect, 201
Dioptric process, 178
Diplopia, 208
Direct action of nervous system, principle, 491
  change in innervation, principle, 491
Directed movements, initiation and control, 109
Discharge, nervous, variations, 393
Discrimination of pitch, 278
    two points on skin, 407
Disparation, 206
Dispersion, measures, 840
Dissonance, 289, 294
Distribution of effort, 584
    and perseveration, 587
  of intelligence, 688
  leptokurtic, 861
  mesokurtic, 860
  platykurtic, 861
Disturbance, propagated, 137
Dominance, group, and leadership, 798
Drainage theory, 559
Drugs, mental effects, 819
Duplicity theory, 255, 267
Dynamics, social, 778

Ear, anatomy, 313
Ebullience, 881
Educational guidance, 680
  tests, correlations, 731
    and evaluation of teaching, 728
      individual diagnosis, 729
      motivation, 729
    origin and development, 708
    and research, 728
      specialization, 730
      surveys, 728
    use, 728
Effect, law, 552, 620
Effector organs, definition, 130
    types, 134
Effort, distribution in learning, 584
Electrical expression of emotion, 514
Elements in visual experience, 170
Eminent men, social background, 759
Emotion, association of analogous sensation, 491
  anatomy and physiology of facial expression, 492
  autonomic control of viscera, 460
    function, 513
  blood changes, 504
    pressure, 503
    sugar, 468, 504

  volume, 502
bodily activity, 501
  change, 449
central mechanisms, 469
cerebrum, 513
connection of movement with sense ideas, 491
criteria, 489
Darwin's principles, 490
in decorticate animals 471
definition, 488
diencephalon as seat, 474
and digestion, 460
direct action of the nervous system, 491
  change in innervation, 491
electrical expression, 514
emergency function of sympathico-adrenal system, 457, 463
endocrine rôle, 510, 512
experimental studies of facial expression, 495
evaluation of knowledge, 488
expression, 488, 519
  conclusions, 519
  future development of research, 520
facial expression, 490
and feelings, 516
gastro-intestinal activity, 508
glycogen, 504
gonads, 511
heart reactions, 505
individual differences, 749
in infants, 649
interpretation of facial expression, 493
James-Lange theory, 478
  tests, 479
James's principles, 492
metabolic rate, 507
methods of study, 488
neuro-humoral basis, 449
peripheral nervous system, 512
pituitary body, 511
principle of antithesis, 490
relation to facial expression, 500
respiratory reactions, 505
rôle of nervous system, 512
as serviceably associated habits, 490
spinal cord, 512
subcortical base, clinical evidence, 475
suprarenal glands, 510
and thalamus, 476, 512
theories, 478
theory based on diencephalic processes, 483
thyroid gland, 511
  visceral nerves concerned, 450
    responses, 502
  vocal expression, 490, 497
  Wundt's principles, 490
Emotional behavior, maturational aspect, 656
  paralysis, 477
Encephalization, 161
Endocrine glands and expression of emotion, 510
  types, heredity, 25
Engram, localization, 540
Entoptic phenomena, 205
Environment, differential effects, 743
  influences in tests, 695
Equipment for observation of infant behavior, 638
Error, probable, 845
  standard, 842
  in statistical data, 833
Errors of sampling, 842
Erscheinungsweisen, 173
Eustachian canal, 315
Evolution, neuromuscular, 128
  of vestibular apparatus, 426
Excitation, 136
  a definition, 352
  and sensation, 350
Excitatory agent, a definition, 352
Exercise, law, learning, 551, 620
Expression of emotion, conclusions, 519
  definition, 488
Extreme cases, special classes, 679
Eye, anatomy and physiology, 177
  characteristics, 174
  cross-section, 176
  movement and motion impression, 210

Facial expression, anatomy and physiology, 492
    in emotion, historical, 490
    experimental studies, 495
    interpretation, 493
    relation to emotion, 500
Fatigue, auditory, 305
    absolute threshold, 307
  muscular, 143
  nerve, 143
  reflex arc, 153
Fechner function, 184
Feeble-minded, classes, 810
  definition, 810
  social background, 759
  and thyroid, 813
Feeling, correlated with variations in nerve discharge, 394
  definition, 488
  and emotion, 516
  qualitative theories, 399

  quantitative theory, 395
  sense, 392
    adaptation, 408
    affective qualities, 410
    learned responses, 406
    localization, 400, 406
    stimuli, 392
Felt experience, 392
  movement, 408
Fertilization, 12
Film color, 173
Flexion reflex, 149
Flicker, 187
  binocular, 204
  chromatic, 200
  photometer, 200
Flight of ideas, 827
Forced movements, 48
Forgetting, rate, 605
Form, visual, distortions, 205
    monocular laws, 204
Fourier's law, 274
Fovea centralis, 177
Frequency, law, 620
  in photoreceptive process, 262
  polygon, 836
Function, central nervous, 146
  Fechner, 184
Fusion and rivalry, 203

g factor in intelligence, 687, 880
Galvanotropism, 67
Gap, tonal, 279
Gastro-intestinal activity in emotion, 508
Gene, 8
  new definition, 83
General ability, 661
    tests, summary, 699
  intelligence, concept, 685
    definitions, 686
    Spearman's view, 687
    theory and practice in test, 688
    Thorndike's view, 686, 878
  norms, 727
  paralysis of the insane, 819
Genius, definition, 810
Geographical series, 833
Geometric mean, 839
Geotropism, 70
Germ-cells, maturation, 7
Gestation, human, duration, 630
Glow, 173
Glycogen in emotion, 504
Gonads in emotional expressional, 511
Grade norms, 727
  scores, 724
Graphs and tables, statistical, 835
Group dominance and leadership, 798
  factors in intelligence, 880
  subject, 802
  tests, 662
    contest, 669
    evaluation of scores, 675
    level, 674
    list, 674
Growth of brilliance, 186
  curves, 836
  factors influencing, 526
  of intelligence, 530, 690
      limit, 692
      regularity, 693
Guessing, correction for in scoring, 726
Guidance, educational, 680
Gustatory papillae and retinal elements compared, 384
  stimuli, 372

Habit interference, 611
  vs. memory, 524
  organization, characteristics, 542
    nature of reaction, 544
        stimulus, 542
  peripheral *vs.* central processes, 549
  (See also *Learning*)
Handicapped children, tests, 684
Harmonic mean, 839
Hearing, 273, 313
  Boring's hypothesis, 339
  limits, 274
  physiology, 313
  resonance hypothesis, 320, 340
      criticism and replies, 340
  sound-pattern theories, 318
  sound-wave theories, 319
  telephone theories, 318
  theories, 318
    and animal experiments, 323
    pathological evidence, 322
    physical evidence, 324
    physiological evidence, 320
  volume hypothesis, 319
  Watt's hypothesis, 335
Heart reactions in emotion, 505
Heat-grill, 403
Hebephrenia, 822
Hereditary Genius, 27
Heredity of acquired characters, 1
        physiological evidence, 3
        psychological evidence, 2
        zoölogical evidence, 4
  action of environment, 5
  all-or-none principle, 9
  chromosome mechanism, 8
  crossing over, 20
  human, and psychology, 22
  in identical twins, 28

  mechanism and laws, 1
  and memory, 1
  Mendelian, 7
  modifiers of characters, 18
  multiple-factor, 18
  of neural patterns, 525
  principles of, Galton, 27
    Pearson, 27
    Yule, 27
  unitary effect, 83
Heterogeneity of maturity, sex and race as general factor in intelligence, 880
Higher mental processes, individual differences, 750
Histogram, 835
Horopter, 208
Hue, 171
  extra-spectral, 192
  function of wave-length, 189
Human behavior, nature and nurture, 32
  communities, integration, 777
  heredity and psychology, 22
Hunger *vs.* appetite, 434
  as automatic stimulus, 434
  and behavior, 117
  as general sensation, 435
  nature and basis, 435
  pangs, local origin, 436
  theories, 435
  and thirst, 434
Hysteria, 827

Identical twins, 28
    maturational correspondence, 651
Idioretinal light, 183
Idiot, 810
Illusions, optical, 206
  stroboscopic, 210
Imbecile, 811
Immobility, tonic, 118
Incest avoidance as culture pattern, 793
Individual differences, 661
    anatomical, social significance, 750
    in animal societies, 745
    intelligence, social significance, 758
    list of traits, 748
    mechanism, 31
    personality, social significance, 766
    physiological, social significance, 754
    psychological, social significance, 757
Individual psychology, definition, 739
    and experimental, 738
  reaction in culture contact, 804
    tests of intelligence, 662
Individuality of infant, 658
Inertia, general factor in intelligence, 880
Infancy, 628
  concepts, 628
  developmental zero, 631
Infant, adaptive and learning behavior, 648
  behavior, range, 642
  emotion and personality, 649
  individuality, 658
  language behavior, 646
  limitations of conditioning, 657
  motor development, 646
  perceptual behavior, 645
  supportive chair, 638
Inheritance of intelligence, 697
  (See also *Heredity*)
Inhibition, drainage theory, 559
  reflex, 154
  retroactive, 599
Injury current, 137
Inner ear, 316, 421
Inquiries into Human Faculty, 34
Insects, taste in, 384
Insight, nature, 575
Inspiration-expiration ratio, 506
Instinct, delayed, 529
  individual differences, 749
  maturation, 529
Integral social psychology, 769
Integration, 160
  in animal societies, 774
    higher organisms, 771
    human communities, 777
    lower organisms, 769
  social, and communication, 780
      social control, 782
Intelligence and citizenship, 763
  cleverness a general factor, 880
  correlation in relatives, 697
  of dependents, 762
  distribution, 688
  factor theory, 880
  family histories, 698
  general, concept, 685
    definitions, 686
    individual differences, 750
    Spearman's view, 687, 879
    Thorndike's view, 686, 878
    two factor theory, 879
  growth, 530, 690
    limit, 692
    regularity, 693
  heterogeneity of maturity, sex and race as general factor, 880
  independent group factors, 880
  individual differences, social significance, 758
  inertia a general factor, 880

  inheritance, 697
  and occupational levels, 760
  perseveration a general factor, 880
  persistence a general factor, 880
  and reflex conduction, 530
  and social background, correlation, 762
    social progress, 764
    social status, 698
  superior children, 698
  tests (see *Tests*)
Intensity, brilliance as function, 183
  chroma as function, 197
  difference limen, 282
  discrimination, 244
    and ocular elements, 261
  law, learning, 555, 620
  maximum for audition, 281
  minimum for audition, 281
  (See also *Brilliance*)
Interference of habits, 611
Intertone, 290
Intra-ocular modification, 183
I. Q. in sectioning, 678
Irradiation, 559
  of reflexes, 161
Irreversibility, 152
Island, tonal, 279

James-Lange theory, 478
    tests, 479
James's principles of emotion, 491

Kelvin scale, 96
Kinaesthesis, 409
Kirschmann's law, 200
Kundt's rule, 240
Kurtosis, coefficient, 860

Labyrinth, anatomy, 421
  function, 422
  and space, 414
Language behavior of infants, 646
Latent period, photosensory, 224
Laughter and crying, 498
Law of segregation, 16
    specific nerve energy, 397
Laws of heredity, Mendel's, 8
Leadership and group dominance, 798
Learning and age, 588
  age of stimulus method, 565
  capacity, phylogenetic development, 531
  central integration, 545
  and chemical changes at cell contact, 557
  curve, 567
    plateaus, 569, 572
    type, 836
    Vincent, 567
  definition, 564
  distribution of effort, 584
      and perseveration, 587
  drainage theory, 559
  experimental studies, 564
  finer neural structure, 541
  and diffuse functional correlations, 558
    growth of new neural processes, 556
    increase in conducting substance, 557
    persistence of excitatory processes, 558
  habit interference, 611
  in infants, 648
  initial and final attainments, 614
  irradiation theory, 559
  insight, 575
  law of contiguity, 551
      effect, 552, 620
      exercise, 551, 620
      frequency, 620
      intensity, 555, 620
      primacy, 554
      recency, 555, 620
      vividness, 555
  laws, 551, 620
  limits, 536
  and maturation in twins, 654
  mechanisms, conclusions, 561
  methods of study, 564
  neural mechanisms, 524
    structure, 537
  paired associates method, 565
  phenomena, 550
  and quantity of neural tissue, 537
  rate, 534
  recognition method, 565
  reinstatement method, 565
  relation of neural structures, 537
  reminiscence, 604
  responses in feeling senses, 406
  retroactive inhibition, 599
  savings method, 565
  speed and retention, 594
  studies, general character, 564
  theories of mechanism, 556
  theory of completeness of response, 621
  transfer of training, 611
  trial and error, 560
  and tropistic behavior, 58
  whole *vs.* part, 535, 581
  (See also *Habit*)
Least squares, method, 862
Leptokurtic distribution, 861
Level of group tests, 674
  occupational and intelligence, 670
Life-cycle and infancy, 628

Light, 181
  adaptation, 187, 229
Limen, intensity difference, 282
Limits of training, 536
Line of regard, 179
Living organisms, study, 45
Local sign doctrine, 400
Localization, auditory, 295
  of engram, 540
  in feeling sense, 400, 406
  of neural function, 545
Long-circuiting in conduction, 161
Loudness, 279
Lumen, 175
Lumens per watt, 183
Luster, 173

Macula lutea, 178
Mania, 827
Manic-depressive psychoses, 826
Masking of tones, 287
Maturation of germ-cells, 7
    instinct, 529
  and learning in twins, 654
    physical handicap in infants, 655
Maturational aspect of emotional behavior, 656
  correspondence in identical twins, 651
Maturity, sex and race, in general intelligence, 880
Mean deviation, 840
  square contingency, 853
Measures of central tendency, 837
    dispersion, 840
    reliability, 869
    special abilities, 705
    validity, 869
Measurement (see *Tests*)
Meatus, 314
Mechanisms, central, of emotion, 469
  of learning, theories, 556
  nervous, in learning, 524
  of reaction, summary, 162
Mechanoreceptors, 352
Median, 837
Melancholia, 826
Membrane theory of conduction, 138
Memory curves, 609
  vs. habit, 524
  and heredity, 1
  (See also *Learning*)
Mendelian heredity, 7
  sex-linked character, dichromatism, 202
Mental defectiveness, heredity, 25
  disease, brain-spot hypotheses, 814
    functional, 815
    mind-twist hypotheses, 814
    and old age, 816
    pathological, 815
    syphilitic infection, 819
Mesokurtic distribution, 861
Metabolism in nerve, 143
  rate in emotion, 507
Method of least squares, 862
Mid-brain in emotion, 512
Middle ear, 315
Mind-twist hypotheses, 814
Mode, 838
Modes of color appearance, 173
Modifiability of behavior, 87
Modifiers of characters in heredity, 18
Monochromatism, 203
Monocular criteria of depth, 206
Moron, 812
Motion impression and eye movement, 210
  perception, thresholds, 209
  visual perception, 209
Motor activity, gradation, 159
  development in infants, 646
  variables, individual differences, 749
Moving average method in smoothing curve, 836
Movement, circus, 55
  directed, initiation and control, 109
  felt, 408
  and position, 409
  random, 113
  spontaneous, 113
Multiple correlation, 857
  factor in heredity, 18
Muscle, fatigue, 143
  contraction, 144
Musical scale, 294

Nasal cavities, air movements, 361
    anatomy, 355
  fossae, cross-section, 355
Natural Inheritance, 27
Nature and nurture in human behavior, 32
Negative after-images, 188
Nerve, fatigue, 143
  impulse, 137
    discharge, variation, 393
    identity, 266
  metabolism, 143
  net, 132
    in higher forms, 134
  trunk *vs.* reflex conduction, 157
Nerves, visceral, in emotion, 450
Nervous conduction, properties, 513
  mechanisms in learning, 524
  system, autonomic, 451
      in emotion, 513
    centralized *vs.* diffuse, 130
    in emotion, 512
    injuries and learning, 545

and learning, 537
nerve-net *vs.* synaptic, 132
parasympathetic, 451, 458
sympathetic, 451, 452
synaptic, origin, 533
Neural function, general, 545
specialization, 545
patterns, heredity, 525
processes, growth, and learning, 556
structure and learning, 537, 541
tissue, quantity, and learning, 537
Neurasthenia, 829
Neurin, 559
Neurofibrils, function, Parker's theory, 354
Neuro-humoral basis of emotion, 449
Neuromuscular evolution, 128
Neurons, fundamental structure, 134
Neurophysiology of behavior patterns, 635
Neurosis, definition, 810
Noise, 283
and tone, 283
Normal probability distribution, 836
Norms, age, 726
for comparison of tests, 714
general, 727
grade, 727
special, 727
Nurture and nature in human behavior, 32
Nystagmus, 423
measurement, 431

Objectivity of scoring in tests, 713
Observation errors, 834
Occupational levels, and intelligence, 760
Ocular elements, number, 259
Odors, classification, 365
Ogive curve, 836
Olfacties, 362
Olfaction, 355
adaptation, 363
Olfactometer, 362
Olfactory cell, 357
receptors, 355
One-way vision screen, 641
Optical defects, 179
illusions, 206
Orientation, photic, quantitative description, 61
in space, 414

Pain sense, 405
Paired associates method, learning, 565
Papillae, 369
Paracusis, 305
Paranoia, 824
Paranoid dementia precox, 824
Parasympathetic nervous system, 451, 458
Partial coefficients, tables, 859
correlation, 855
Pathological evidence in hearing theory, 322
Patterns of behavior, 633
(See also *Behavior Patterns*)
of discharge in sense of feeling, 396
Percentile scores, 724
Perception and sensation in vision, 170
space, auditory, 295
questions, 415
tone and number of vibrations, 279
vestibular data, characteristics, 417
Perceptual behavior, infant, 645
Performance scales, 666
Periods of human development, 628
Peripheral nervous system in emotion, 512
processes in learning, 549
Perseveration and distribution of effort, 587
general factor in intelligence, 880
Persistence a general factor in intelligence, 880
of vibrations, 329
Personality, definitions, 766
individual differences, social significance, 766
of infants, 649
Phase, definition, 276
relative, in hearing, 328
Phenomena of vision and stimulus correlations, 169
Phi gamma function, 836
phenomenon, 210
Photic orientation, quantitative description, 61
Photochromatic interval, 197
Photographic observation dome, 641
Photometer, flicker, 200
Photometric measures, 175
Photon, 175
Photopic adaptation, 187
Photoreceptor process, chemical substances, 238
nature, 216
physicochemical nature, 236
Phototropic effect, measurement, 55
Phototropism, 49
Phylogenetic development of learning capacity, 531
Physical criterion of pitch, 286
handicap and maturation in infants, 655
models in auditory theory, 324
structure, individual differences, 748

Physics and auditory theory, 324
    behavior, 45
  of sound, 273
Physiognomy, individual differences, social significance, 752
Physiological foundations of behavior, 770
  gradient, 770
Physiology, individual differences, 749
    social significance, 754
Physique, individual differences, social significance, 750
Pinna, 314
Pitch, 277
  discrimination, 278
  physical criterion, 286
  of pure and compound tones, 286
Pituitary body in emotional expression, 511
Plateaus, 569, 572
Platykurtic distribution, 861
Pore plates, 360
Portomento, 346
Position and movement, 409
Positive after-images in brilliance decay, 186
Practice in intelligence tests, 694
Prenatal behavior, 644
  stage, periods, 632
Primacy, law, 554
Primitive people, psychology, value, 786
Primitiveness of emotional reactions, 470
Principle of visual eclipse, 174
Principles of heredity, Galton, 27
    Pearson, 27
    Yule, 27
Probable error, 845
Product-moment correlations, 847
  scales, 720
Progress, social, and intelligence, 764
Propagated disturbance, 137
Proprioceptive reflex, 151
Protanopia, 202
Psychasthenia, 829
Psychogalvanic reflex, 514
Psychological color solid, 171
  variables, individual differences, 749
  variation, social significance, 757
Psychometry, 635
Psychoneuroses, 827
  hysteria, 827
  neurasthenia, 829
  psychasthenia, 829
Psychophysical methods, mathematical principles, 866
Psychophysics of brilliance, 179
    chromatic attributes of visual phenomena, 189
Psychosis, definition, 810
Puberty precox, 655
Pulsation of tone, critical frequency, 326
Pupil, eye, 179
Pure tone sensations, number, 282
Purkinje phenomenon, 181
Purples, 192

Qualitative series, 833
  theories of feeling, 395
Qualities of felt experience, 400
  taste, 370
  of vestibular sensations, 417
Quantitative series, 833
  theory of feeling, 395
Quartile deviation, 840
Quotients as scores, 725

Race differences, anatomical, 753
    in tests, 682
Radiant energy, stimulus in vision, 174
Radioreceptors, 353
Random movements, 113
Rate of learning, 534
Raw scores, 724
Reaction of cultures, 794
  mechanism, 128
    summary, 162
  time, individual differences, social significance, 757
Recall method, learning, 565
Recency, law, 554, 620
Recept, 350
Receptor, fundamental property, 352
  organs, definition, 130
  olfactory, 355
  for sense of feeling, 392
Reciprocal innervation, 149
Reciprocity law, 186
Recitation and retention, 596
Recognition method, learning, 565
Recruitment, 151
Rectilinear correlation, 846
Reflex arc, simplest, 132
  conduction, characteristics, 147
    and intelligence, 530
  cross extension, 149
  flexion, 149
  inhibition, 154
  irradiation, 161
  vs. nerve trunk conduction, hypothesis, 157
  proprioceptive, 151
  spinal, 148
Refractory phase, 139
Regression, 847
Reinforcement in reflex arcs, 154

Reinstatement method, learning, 565
Reissner's membrane, 316
Relationship, measures, 845
Reliability coefficient, 870
  measures, 869
  of tests, 714
Reminiscence, 604, 608
Research, use of tests, 728
Resistance and weight, 410
Resonance hypothesis of hearing, 320, 340
  sharpness, 330
  theory, criticisms and replies, 332
Resonators, properties, summary, 332
  relationship between properties, 329
Respiratory reactions in emotion, 505
Retention, effect of active and passive recall, 596
  influence of activity, 599
  loss through disuse, 603
  and recitation, 596
    speed of learning, 594
Retina, cortical, 546
Retinal elements, see ocular elements
    and gustatory papillae compared, 384
Retroactive inhibition, 599
Reversion toward mediocrity, Galton, 27
Rheobase, 136
Rivalry of culture groups, 798
  and fusion, 203
Rods and cones, 177
Roscoe-Bunsen law, 57, 218
Rotation, stimulus in vestibular sense, 418

s factor in intelligence, 687
Salt taste, stimulus, 373
Sampling errors, 834
  theory, 880
Saturation, 171
  as function of wave-length, 189
Savings method, learning, 565
Scale, musical, 294
Scales, Binet, 663
  performance, 666
Scaling tests, 721
Scatter-diagram, 845
Science, definition, 45
Score, T, 725
Scores, age and grade, 724
  correction for guessing, 726
  group tests, evaluation, 675
  quotients as, 725
  raw 724
  standard, 724
  types, 724
  weighted, 726
Scoring, speed *vs.* accuracy, 723
    as basis, 722
  tests, 722
    difficulty of task as factor, 723
    quality of performance, 723
Scotopic adaptation, 188
Sectioning, use of I. Q., 678
Semantic aphasia, 548
Semi-interquartile range, 840
Senility and mental disease, 816
Sensation, a definition, 351
  and excitation, 350
    perception in vision, 170
Sense-organs, common characteristics, 217
  dual function, 217
  similarities, 216
  threshold, 218
Senses, chemical, 350
Sensitivity, skin, 407
  visceral, 410
Sensitization period, 222
Sensory acuity, individual differences, 749
  appropriation, 353
  process, duality, 222
  system as buffer, 269
Serviceably associated habits, principle, 490
Sex distinction as culture pattern, 793
  social significance, 754
Sharpness of resonance, 330
Sherrington's hypothesis of conduction, 157
Short-answer tests, 719
Simple functions, tests, 661
  reflexes and emotion, 470
Size and brilliance, 185
Skewness, coefficient, 860
Skin, acuity, 408
  sensitivity, 409
Smell, 350, 355
Smoothing curves, method of moving averages, 836
Social background of eminent men, 759
    and intelligence correlation, 762
  behavior, primitive, 791
  control, 782
  dynamics, 778
  life, community affair, 790
  progress and intelligence, 764
  psychology, 738, 786
    conclusions, 783
    definition, 739
    differential, 742
    integral, 769
    synthesis of views, 740
  status and intelligence, 698
Sociological views of society, 787

Somatic induction *vs.* direct action of environment, 6
Sound, general description, 273
  pattern theories of hearing, 318
  physics, 273
  wave theories of hearing, 319
  waves, compound, 274
    reflection, 275
    refraction, 275
    simple, 274
Sour, stimulus, 374
Space, auditory perception, 295
  and non-auditory labyrinth, 414
  orientation, 414
  perception, questions, 415
  and vestibular data, 416
Spacial determinants of chroma, 198
  factors, brilliance as function, 185
Spearman's factor theory of intelligence, 687, 880
Special abilities, constitution, 729
    measurement, 705
  classes, 779
  norms, 727
Specialization and educational tests, 730
  of neural function, 545
Specific energy of nerves, 266, 397
Spectral chroma scale, 189, 190
Speech, 303
  as culture pattern, 793
Speed *vs.* accuracy in scoring tests, 723
  as basis of scoring tests, 722
  of conduction in coordination, 160
    learning, 534
      and retention, 594
Spinal cord, composition, 134
    in emotion, 512
  reflexes, 148
Spontaneous movements, 113
Spots of sensitivity, skin, 401
Square contingency, 853
Standard deviation, 841
  error, 842
  scores, 724
  tests, types 718
Standardization of tests, characteristics, 710, 718
      form of test questions, 712
      method of administration, 713
      methods, 718
      norms for comparison, 714
      objectivity of scoring, 713
      reliability, 714
      selection of samples of performance, 710
      validity, 716
Stationary state, 248
Statistical data, errors, 833
  method, purpose and function, 832
  series, geographical, 833
    qualitative, 833
    quantitative, 833
    temporal, 833
  treatment of typical problems, 855
Statistics, correlation, 846
  curve-fitting, 859
  errors of sampling, 842
  general treatment, 832, 855
  measures of central tendency, 840
      dispersion, 840
      relationship, 845
      reliability and validity, 869
  multiple correlation, 857
  partial correlation, 855
  psychophysical methods, 866
  tables and graphs, 835
Stereotropism, 49
Stimuli for sense of feeling, 392
  vestibular, nature, 417
Stimulus attributes, visual, 177
  correlations of visual phenomena, 169
  a definition, 352
  time element, 392
Stroboscopic illusion, 210
Student, college, tests, 680
Subject group in cultural adjustment, 802
Subliminal stimulation, 393
Summation, 152, 393
Sunlight, spectral energy, 176
Superior children, relatives, 698
Suprarenal glands in emotion, 510
Surface color, 173
Surveys, educational, 728
Survival of cultures, 806
Sweet taste, stimuli, 379
Sympathetic nervous system, 451
Sympathico-adrenal system, emergency function, 457, 463
Synaptic nervous system, origin, 132, 533
Syntactical aphasia, 548
Systematic errors of observation, 834

T score, 725
Tables of partial coefficients, 859
  and graphs, statistical, 835
Tactile and auditory sensations, 280
Talbot law, 57
  -Plateau law, 187
Taste, 350, 367
  buds, 368
    distribution, 372
  qualities, 370
    stimuli, 372
Telephone hypothesis, 318
  theories of hearing, 318

Temperature and behavior, 55
  coefficient, 221
  sensitivity, 402
Temporal series, 833
Test, CAVD, 687
  score, T, 725
    age, 724
    correction for guessing, 726
    grade, 724
    percentile, 724
    quotients, 725
    raw, 724
    standard, 724
    weighted, 726
Testing, intelligence, uses and results, 677
  some statistical problems, 874
Tests, accuracy *vs.* speed, 723
  Binet scales, 663
  blind, 685
  of college students, 681
  deaf, 684
  difficulty of task in scoring, 723
  educational, origin and development, 708
    use, 728
  general ability, summary, 699
    and analytical, 720
  group, content, 669
    evaluation of scores, 675
    level, 674
    list, 674
  handicapped children, 684
  intelligence, development, 661
    general environmental influences, 695
    group and individual, 662
    performance scales, 666
    practice and coaching, 694
    summary, 699
    theory and practice, 688
  limitations as measures of ability, 732
  product scales, 720
  quality of performance in scoring, 723
  race differences, 682
  scaling, 721
  scoring problems, 722
  short-answer, 719
  special aptitude, origin and development, 706
  speed as basis of scoring, 722
  standard, types, 718
  standardization, characteristics, 710
    form of test question, 712
    method of administration, 713
    methods, 718
    norms for comparison, 714
    objectivity of scoring, 713
    reliability, 714
    selection of samples of performance, 710
    validity, 716
  types of scores, 724
  unit-task, 718
  (See also *Educational Tests*)
Thalamus and emotion, 476, 512
Theory of sampling, 880
Thirst *vs.* appetite, 434
  as automatic stimulus, 434
  evidence of local origin, 444
  and hunger, 434
  nature and basis, 440
  theories, 441
Thorndike's view of general intelligence, 686, 878
Three sensation curves, 266
Threshold, absolute, for brilliance, 183
  of sense-organs, 218
Thyroid and feeble-mindedness, 813
  gland in emotional expression, 511
Time, brilliance as function, 186
  chroma as function, 200
  of conduction, 515
Tonal islands and gaps, 279
  perception and number of vibrations, 279
Tone character, 277
  critical frequency of pulsation, 326
  musical, half wave-length change, 326
  and noise, 283
  pitch of pure and compound, 286
  sensations, pure, number, 282
Tones, combination, 289
  compound, 283
  masking, 287
  pure, comparative loudness, 288
  simple, 276
Tonic immobility, 118
Tonus, 159
Transfer of training, 611
Tremulo, 346
Trial and error learning, 560
Trichromatism, 202
Trireceptor theory, 266, 267
Tropism, definition, 54
  historical, 48
  as instrument of analysis, 54
Tropisms, 48
  modification, 53
Tropistic behavior, compounding, 67
    and learning, 58
    problems, 54
Twins, identical, and heredity, 28
    maturational correspondence, 651
  maturation and learning, 654
Two factor theory of intelligence, 879

point discrimination, 407
Tympanic membrane, 314

Unit-task tests, 718
Universal response, culture determinant, 792
Upset, definition, 488

Validity, coefficient, 869
  measures, 869
  of tests, 716
Variability of behavior, 87
    threshold in reflex arc, 153
Variables in individuals, table, 748
Vegetative nervous system, 134
Vestibular apparatus, phylogenesis, 425
  data in animals, 426
    human, 426
    and space perception, 416
  perceptions, characteristics, 417
  research, technique and method, 428
  sensations, 416
  stimulation, non-perceptual, 423
    and viscera, 425
  stimuli, nature 417
Vestibule, galvanic stimulation, 428
Vibratory sense, 410
Viscera, autonomic control in emotion, 460
Visceral responses in emotion, 502
  sensitivity, 410
Visibility coefficient, 181
  curves, 180
  values, average normal, 182
Vision and the cortex, 211
  the dioptric process, 178
  phenomena of, 169
  theory, 216
Visual acuity, function of stimuli characteristics, 204
    and illumination, 253
    number of ocular elements, 259
  discrimination, intensity, 244
  eclipse, principle, 174
  field, subjective, 169
  form, depth and the binocular determination, 206
  monocular, laws, 204
  perception of motion, 209
  phenomena and stimulus correlations, 169
  process, 216
  purple, as photosensitive substance, 240
  space, general properties, 174
    theory, 208
  stimulus attributes, 177
   characteristics, 174
Vividness, law, learning, 555
Vocal expression of emotion, 497
Vocalization in emotion, 490
Volume, auditory, 301
  hypothesis of hearing, 319

w factor in intelligence, 880
Watt's hypothesis of hearing, 335
Wave-length, brilliance as function, 180, 185
  chroma as function, 191
  color excitations as function, 196
  hue as function, 189
  saturation as function, 189
Weber-Fechner law, 55, 72, 245, 252, 417
Wedensky effect, 155
Weight, sense, 410
Weighted scores, 726
Whole *vs.* part learning, 535, 581
Wundt's principles of emotion, 490

Zero, developmental, 631
Zones, chromatic, 199

University of British Columbia Library

DUE DATE

NOV 14 1972
NOV 28 1972
DEC 13 1972
DEC 13 1972
DEC 18 1972 REC'D
APR 08 1980
APR 8 1980 REC'D
AUG 26 1986 REC'D

FORM 310

UNIVERSITY OF B.C. LIBRARY
3 9424 01062 4242